Regulation of POU transcription factor activity by OBF1 and Sox2

Thesis

Regulation of POU transcription factor activity by OBF1 and Sox2

(44,000 words)

A thesis submitted in partial fulfillment of the requirements of the

Open University

for the degree of Doctor of Philosophy

by

Katharina Lins

Sponsoring Establishment: NIMR, Mill Hill, London

Collaborating Establishment: Center for Animal Transgenesis and

Germ Cell Research, University of Pennsylvania,

New Bolton Center, Kennett Square, PA 19348, USA

Karlsruhe, April 2004

To my parents

and Julia

Contents

Declaration

This dissertation is the result of my own work. It contains nothing, which is the outcome of work done in collaboration, except where clearly stated as such.

The work, or any part of it, has not previously been submitted for any other degree or qualification.

Katharina Lins

April 2004

Acknowledgments

I would like to thank Dr Hans Schöler for giving me the opportunity to work on my PhD at the EMBL in Heidelberg and the University of Pennsylvania. I am very grateful for his advice, support and especially his encouragement. He was always excited about my projects even when I got fed up with them.

Furthermore, I would like to thank Dr Robin Lovell-Badge, my secondary supervisor at the NIMR in London and Dr Austin Smith for their input and assisting me with the organizational aspects of my thesis and viva.

I would also like to thank Dr Matthias Wilmanns who gave me the opportunity to work on a project at the EMBL in Hamburg and Dr Patrick Matthias and Steffen Massa for their collaboration.

Moreover, I would like to thank all members of the Schöler and Wilmanns labs who assisted me in one way or another on my journey through the lab jungle. I am especially grateful to Dr Valérie Botquin, who taught me all there is to know about molecular biology. Furthermore, I enjoyed Drs Attila Reményi's and Alexey Tomilin's close collaboration and intensive discussions.

Last but not least I want to thank Dr Stefan Schlatt, Anna and Julius Velde and my families in Gum Tree and Europe for moral support and great input.

Abbreviations

'	minutes
''	seconds
aa	amino acid
APS	ammonium persulphate
ATP	adenosine 5'-triphosphate
β-ME	β-mercaptoethanol
bp	base pairs
BSA	bovine serum albumine
°C	degrees Celsius
CIP	calf intestinal phosphatase
cpm	counts per minutes
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytosine 5'-triphosphate
dGTP	deoxyguanosine 5'-triphosphate
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside 5'-triphosphates
DTT	dithiothreitol
dTTP	deoxythymidine 5'-triphosphate)
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EtBr	ethidium bromide
FCS	fetal calf serum

g	grams
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hrs	hours
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase
L-Glu	L-Glutamine
M	molar
mA	milli Amperes
mCi	milli Curie
μCi	micro Curie
mg	milligrams
μg	micrograms
ml	milliliters
μl	microliters
mM	millimolar
μM	micromolar
mRNA	messenger ribonucleic acid
ng	nanograms
nm	nanometers
OD	optical density
ONPG	O-Nitrophenyl-β-D-galactopyranoside
PAGE	polyacrylamide gel elctrophoresis
PCR	polymerase chain reaction
pg	picograms
pM	picomolar

poly[d(I-C)]	poly deoxycytosine-deoxyinosine
P/S	penicillin/streptomycin
RNA	ribonucleic acid
Rnase A	ribonucleic acid endonuclease A
rcf	relative centrifugal force, also referred to as g-force
rpm	revolutions per minute
RT	room temperature
SDS	sodium dodecasulphate
TEMED	tetramethylethylenediamine
UV	ultraviolet
V	Volts
v/v	volume/volume
w/v	weight/volume

Nucleic acids

A	adenosine
C	cytosine
G	guanidine
T	thymidine

Amino acids

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

List of figures

Abstract

For a cell to exert a specialized function certain genes have to be expressed, others repressed. Transcription factors, regulating this expression, do not function alone, but are often part of multi-protein complexes. Regulating a single gene with more than one transcription factor is an efficient way to integrate responses to a variety of signals using a limited number of proteins. DNA binding proteins often interact with each other and with non-DNA binding proteins in a specific arrangement. The assembly of these complexes is often highly cooperative and promotes high levels of transcriptional synergy.

The center of my thesis is the family of POU transcription factors. Specifically, I elaborate the interaction within the POU protein family, with members of other transcription factor families and with cofactors. In all cases, the assembly of the correct array of polypeptides on the DNA requires specific protein-protein and protein-DNA interactions.

As an example of POU factors interacting with each other and with a cofactor I investigated the properties of a protein-DNA complex with the B-cell-specific cofactor OBF1 and the Oct1 dimer. Depending on the DNA sequence they bind to, Oct1 dimers are arranged in configurations that are either accessible (PORE sequence) or inaccessible (MORE sequence) to OBF1. In Chapter 3 I show that the expression of *Osteopontin*, which contains a PORE sequence in its enhancer region, depends on the presence of OBF1 in B-cells. OBF1 alleviates DNA sequence requirements of the Oct1 dimer on PORE-related sequences *in vitro*. Furthermore, OBF1 enhances POU dimer-DNA interactions and overrides Oct1 interface mutations, which abolish PORE-mediated dimerization without OBF1. Based on the biochemical data, I propose a novel Oct1 dimer arrangement when OBF1 is bound.

As an example of Oct factors interacting with members of another transcription factor family I studied the interactions of Sox2 with Oct1 and Oct4, respectively. POU and Sox transcription factors exemplify partnerships established between various transcriptional regulators during early embryonic development. The combination of Oct4 and Sox2 on DNA is considered to direct the establishment of the first three lineages in the mammalian embryo.

Although functional cooperativity between key regulator proteins is pivotal for milestone decisions in mammalian development, little is known about the underlying molecular mechanisms. The data in Chapter 4 validate experimental high-resolution structure determination, followed by model building. The study shows that Oct4 and Sox2 are able to dimerize on DNA in distinct conformational arrangements. The binding site characteristics of their target genes are responsible for the correct spatial alignment of the Velcro-like interaction domains on their surface. Interestingly, these surfaces frequently have redundant functions and are instrumental in recruiting various interacting protein partners.

In Chapter 5 I investigated how Sox2 and Oct4 regulate transcription of a target gene. The first intron of *Osteopontin* contains a Sox-binding site and a unique PORE to which Oct4 can either bind as a monomer or a dimer. The study reveals that Sox2-specific repression depends on an upstream Sox site and an intact PORE, although neither the Sox nor the PORE sites are negative elements on their own. A mechanism is being proposed how Sox2 represses Oct4-mediated activation of *Osteopontin*.

Chapter 1

Introduction

1.1 Transcription factors

For a cell to exert a specialized function certain genes have to be expressed, others repressed. The expression of genes is primarily regulated at the level of transcriptional initiation of mRNA synthesis. Thus, transcription factors that control cell- or tissue specific transcription are important developmental regulators. In general, these are sequence-specific DNA-binding proteins whose recognition sites are present in promoters and enhancers. Although there is significant variation in the structures of transcription factors, they all share one property: they have various functional domains or regions, one of which is a DNA-binding domain that recognizes a specific DNA sequence within the enhancer or promoter and binds to it. Various different structural motifs that mediate sequence-specific contacts with DNA were revealed by structural studies and sequence comparisons. In most cases, the DNA-binding motif is a distinct structural domain that has been conserved through evolution among a wide range of species, as far as from prokaryotes to mammals. Transcription factors are grouped into families according to DNA-binding motifs such as zinc-fingers, leucine zippers, helix-turn-helix motifs, and helix-loop-helix motifs, to provide just some examples (Pabo and Sauer, 1992). Variations within these structural motifs provide an additional way to recognize different sequences, thus creating functional specificity.

Regions other than the DNA binding domain of the protein are often not well conserved among members of a transcription factor family. These domains are located amino (N)- or carboxy (C)-terminal of the DNA binding domain and may be involved in oligomerization, interaction with other proteins, transcriptional activation or repression, ligand binding, and response to intracellular signals. Deletion experiments and studies with chimeric proteins have indicated that the functional

domains typically operate independently, implying that transcription factors are modular in structure (Frankel and Kim, 1991). One known exception from this rule is the reported interaction between the DNA-binding POU domain and the C-terminal transactivation domain of Oct4 (Brehm *et al.*, 1997)

Transcription factors do not function alone, but rather in concert with other proteins. Regulating a single gene with a combination of more than one protein is an efficient way to integrate responses to a variety of signals using a limited number of transcription factors resulting in either repression or activation of transcription. The permutations of protein arrangements based on the DNA sequence facilitate the establishment of complex regulatory networks, as they indeed exist in higher eukaryotes. Transcription factors bind DNA and often interact with each other in a specific arrangement. The assembly of these complexes is often highly cooperative and promotes high levels of transcriptional synergy (Carey, 1998). One example of a well-studied gene that is activated by an array of transcription factors is the human interferon-β (IFN-β) gene (Maniatis *et al.*, 1998). Upon viral infection IFN-β is activated by an enhanceosome consisting of three transcription activator complexes (including NF-κB, ATF-2/c-Jun heterodimer, interferon regulatory factors IFR-3 and IFR-7 and further transcription factors) and the architectural protein HMG I(Y) (see section 1.3). An enhanceosome is a higher-order nucleoprotein complex whose specific array of transcription factors is required for efficient recruitment of the transcription machinery to the promoter. Thus, the relative rotational positions of the binding sites with respect to each other are crucial for enhancer activity.

In the work presented in this thesis I have studied the interaction of transcription factors with other members of the same or different families or with a transcriptional cofactor. Cofactors often do not recognize DNA elements directly, but

are tethered by interacting specifically with other transcription factors. In all cases that I will describe, the assembly of the correct array of polypeptides on the DNA requires specific protein-protein and protein-DNA interactions.

1.2 Oct factors

Oct factors are a subfamily of the POU transcription factors. The term POU was derived from the names of the first 4 factors shown to have a particular region of similarity, **P**it1 (a pituitary-specific transcription factor), **O**ct1 (a ubiquitously expressed octamer binding protein first shown to regulate immunoglobulin gene transcription), **O**ct2 (another octamer binding factor expressed in lymphoid cells) and ***unc-86*** (a *C. elegans* protein) (Clerc *et al.*, 1988; Finney *et al.*, 1988; Herr *et al.*, 1988; Ingraham *et al.*, 1988; Sturm *et al.*, 1988). The Oct factors were named after the 8bp DNA element they bind to. This so called octamer motif (ATGCAAAT) was first found to regulate the expression of immunoglobulin genes (Falkner and Zachau, 1984; Parslow *et al.*, 1984).

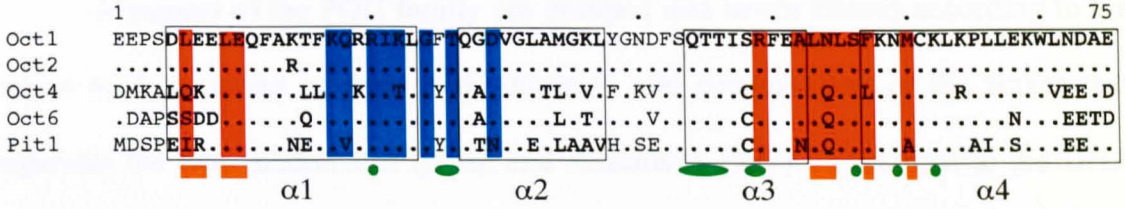
POU factors consist of a DNA binding domain called the POU domain flanked by two transactivation domains (Herr *et al.*, 1988). The POU domain is subdivided into two modular DNA binding subdomains. An unstructured linker region, variable in sequence and length (15-56 amino acids), connects both subdomains (Herr and Cleary, 1995). The C-terminal part of the POU domain consists of a homeodomain, called the POU homeodomain (POU_H), containing the three usual α-helices. The N-terminal domain is unique to the POU proteins and together with POU_H is required for DNA binding. A significant sequence homology with other proteins could not be found so it was called POU specific domain (POU_S). It consists of four α-helices connected by short loops (figure 1.1). Structural comparisons

Figure 1.1

Sequence alignment of POU domains from different transcription factors.

The seven α-helices within the POU domain are indicated. Amino acid residues conserved to Oct1 are indicated by dots. Residues involved in DNA interaction of the monomer are highlighted in green (Klemm *et al.*, 1994), those interacting with OBF1 in orange (Chasman *et al.*, 1999). Amino acids involved in POU_S-POU_H interface formation in the Oct1/MORE and Oct1/PORE dimers are highlighted red and blue, respectively (Reményi *et al.*, 2001). Serine/threonine residue 107, which is an *in vitro* and *in vivo* phosphorylation site (Roberts *et al.*, 1991: Segil *et al.*, 1991; Kapiloff *et al.*, 1991) is marked by a yellow star.

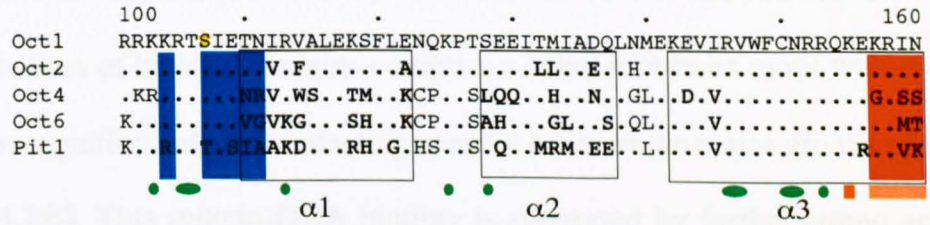
POU specific domains



POU linkers

Oct1	NLSSDSSLSSPSALNSPGIEGLSR
Oct2	TMSVDSSLPSPNQLSSPSLGFDGLPGR
Oct4	NNENLQEICKSETLVQA
Oct6	SSSGSPTNLDKIAAQGR
Pit1	QVGALYNEKVGANER

POU homeodomains



MORE interface

PORE interface

phosphorylation site

monomer interaction with DNA

monomer interaction with OBF1

revealed a high similarity to the prokaryotic bacteriophage lambda and the 434 repressor, despite the lack of sequence homology (Assa-Munt *et al.*, 1993). POU_S favors binding to an ATGC sequence and POU_H recognizes (A/T)AAT resulting in the combined POU recognition sequence of ATGC(T/A)AAAT (Verrijzer *et al.*, 1992).

Members of the POU family are grouped into seven classes according to the amino acid sequence of their POU domains and conservation of the linker that separates the two subdomains (Ryan and Rosenfeld 1997). In contrast to the DNA binding domain, the composition and length of the activation domains differ considerably between various POU proteins, which may partly reflect different activation potential.

Resolution of the crystal structure of the Oct1 POU domain on the octamer site of the H2B promoter by Klemm *et al.* (1994) showed that the second and third helices of both subdomains comprise a helix–turn-helix motif with α-helices 3 (DNA recognition helices) contacting specific bases in the major groove of the DNA (figure 1.2A). This protein-DNA binding is supported by further amino acids that interact with phosphates of the DNA backbone via hydrogen bonds. Almost all residues in the recognition helix of Oct1, which make contacts with DNA, are conserved between POU family members (figure 1.1). When binding the octamer motif, POU_S and POU_H lie on opposite sides of the double helix. There are no protein-protein contacts between the subdomains, but they contact overlapping DNA backbone phosphates. The POU_S and POU_H lie in the major groove of the DNA but the basic N-terminus of POU_H is positioned in the minor groove and makes contacts with the DNA backbone (figure 1.2A). This indicates, that the linker transverses the DNA through the minor

Figure 1.2

Illustration of Oct1 crystal structures solved to date.

(**A**) Overview of the POU1 monomer bound to the octamer motif (Klemm *et al.*, 1994). Upper: numbered cylinders represent the respective α-helices of POU_S and POU_H. The smaller case numbers represent the residues comprising the α-helices. The dotted line shows the estimated path of the flexible linker between the POU subdomains. The octamer motif (ATGCAAAT) is labeled on the DNA helix. Lower: POU_H (red), POU_S (yellow) and the DNA (blue) are illustrated by ribbons. The phosphorylation site serine 107 is pointed out (see section 1.5.4)

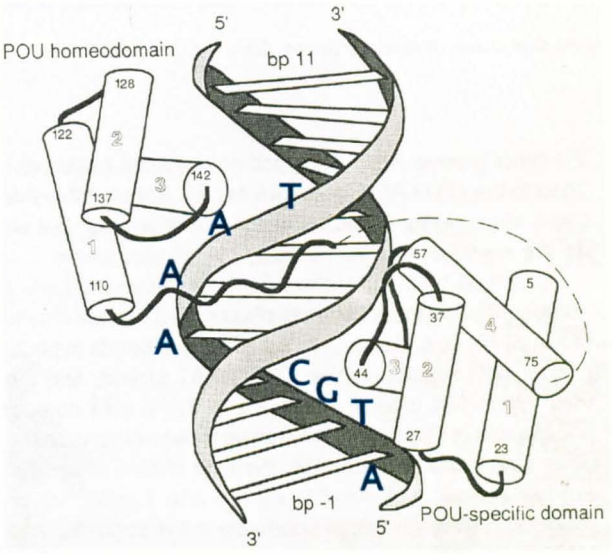
(**B**) Ternary complex consisting of the OBF1 peptide (purple), Oct1 POU domain (POU_H = red, POU_S = yellow), and DNA containing the octamer sequence (Chasman *et al.*, 1999). The OBF1 peptide traverses the major groove at the A-T base pair at position 5 (green) of the octamer site. At the C-terminus the OBF1 peptide forms a short α-helix and makes extensive contacts with POU_S, at the N-terminus it contacts POU_H.

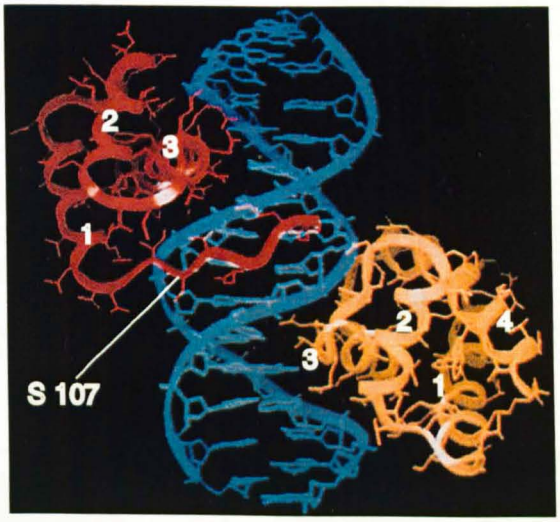
(**C**) POU1 dimerizes on the PORE and the MORE in two distinct configurations (Reményi *et al.*, 2001). The subdomains allocated to the same POU1 molecule are depicted in the same color. N and C designate the N- and C-terminus of each POU molecule, respectively. The dotted line shows the estimated path of the flexible linker between the POU subdomains.

Lower panel: Close-up of the POU_S-POU_H interfaces. On the MORE, I159 from the C-terminal part of POU_H interacts with a hydrophobic pocket consisting of residues from POU_S. The C-terminal helix of the POU_H is depicted in green; the POU_S is shown as a surface representation in purple with the hydrophobic residues in yellow. On the PORE the N-terminal nonhelical part of POU_H is located in the minor groove of the PORE element. One DNA phosphate group forms H bonds with Arg20 (POU_S) and Ser107 (POU_H). POU_S is colored purple; POU_H green. The DNA molecule is shown as a gray surface representation. The image is inverted in comparison to the illustration above it.

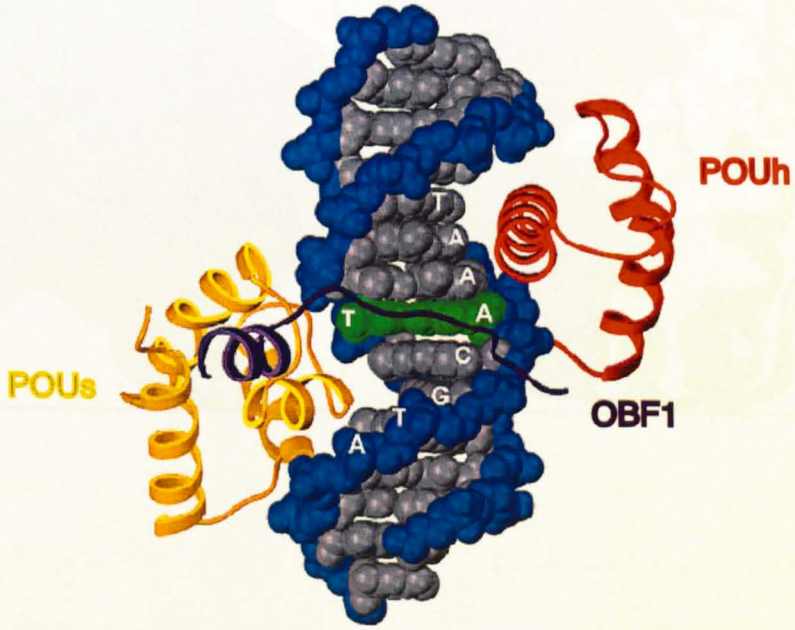
(A, B and C reproduced from Klemm *et al.* (1994); Chasman *et al.* (1999) and Reményi *et al.* (2001), respectively)

A

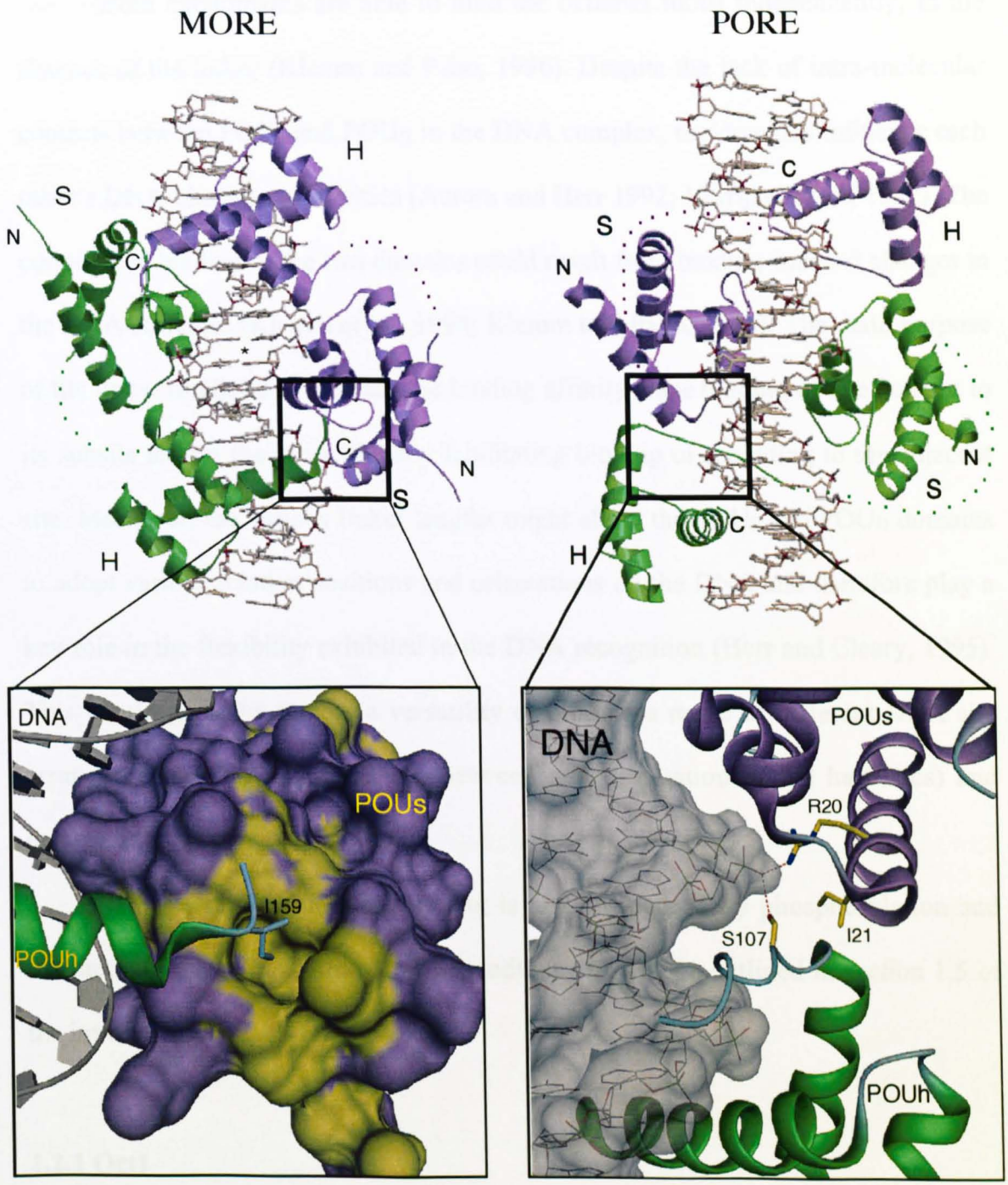

POU homeodomain
5'
3'
bp 11
128
122
2
3
142
137
1
110
T
A
A
A
57
5
37
4
44
3
2
75
1
27
23
C
G
T
A
bp -1
5'
3'
POU-specific domain


2
3
1
S 107
2
4
3
1

B


POUh
POUs
OBF1
T
A
A
T
A
C
G
T
A

C


MORE
PORE
H
S
N
C
C
S
N
H
C
N
S
H
C
S
N
H
DNA
POUs
I159
POUh
DNA
POUs
R20
S107
I21
POUh

groove although it is not detectable in the electron density map of the crystal, revealing that it is distorted in the crystal and not rigid as the subdomains.

Both subdomains are able to bind the octamer motif independently, in the absence of the linker (Klemm and Pabo, 1996). Despite the lack of intra-molecular contacts between POU_S and POU_H in the DNA complex, the domains influence each other's DNA-binding specificities (Aurora and Herr 1992; Verrijzer *et al.*, 1992). The cooperative binding of the two domains could result from binding-induced changes in the DNA structure (Klemm *et al.*, 1994; Klemm and Pabo, 1996). The main purpose of the linker might be to increase the binding affinity since binding of one domain to its subsite tethers the other, thereby facilitating binding or rebinding to the adjacent site. Moreover, the various linker lengths might allow the POU_S and POU_H domains to adopt various relative positions and orientations on the DNA and therefore play a key role in the flexibility exhibited in the DNA recognition (Herr and Cleary, 1995). This would give Oct factors a versatility of binding a range of different DNA site arrangements (varying in spacing between and orientation of the half-sites) and cofactors.

The versatility of DNA binding is also influenced by phosphorylation and interaction on DNA as homo- and heterodimers as will be outlined in section 1.5 of the introduction.

1.2.1 Oct1

Most biochemical studies have been performed with Oct1 serving as a model for the other Oct factors (see sections 1.2, 1.5.1 and 1.5.4). Oct1 is ubiquitously expressed and regulates transcription of small nuclear RNA genes (snRNA) and the histone H2B

gene, which also are ubiquitously expressed genes (LaBella *et al.*, 1988; Segil *et al.*, 1991; Yang *et al.*, 1991; Hinkley and Perry, 1992; Mittal *et al.*, 1996).

In other cases, Oct1 needs to recruit cofactors to upregulate transcription. Oct1 and VP16, for example, activate expression of the herpes simplex virus (HSV) immediate early (HSV IE) genes (Cleary *et al.*, 1993). After viral infection of a cell, VP16 recruits Oct1 to the immediate early genes of the virus. Deletion analyses showed that the VP16 transactivation domain is sufficient for VP16 dependent promoter activation (Sadowski *et al.*, 1988; Cleary *et al.*, 1993). To support sequence specific IE gene activation by its own activation domain VP16 requires the Oct1 POU domain (Wu *et al.*, 1994). The regulatory regions of HSV IE genes contain a POU_H (TAAT) and a 3' POU_S binding site (GARAT; R= purine: A or G). In this case POU_S would bind to the GARAT sequence together with VP16 (Cleary *et al.*, 1995). With the strong VP16 transactivator the herpes simplex virus has acquired an adaptor protein to boost the activity of Oct1. This interplay gives insights into transcription factor-cofactor interaction.

By interacting with cell specific transcription factors or cofactors, Oct1 can also activate genes whose expression is restricted to specific cell types. Together with OBF1, a lymphoid specific cofactor, which I will cover in greater detail in section 1.4 of the introduction and Chapter 3, Oct1 can upregulate lymphoid-specific genes (figure 1.3).

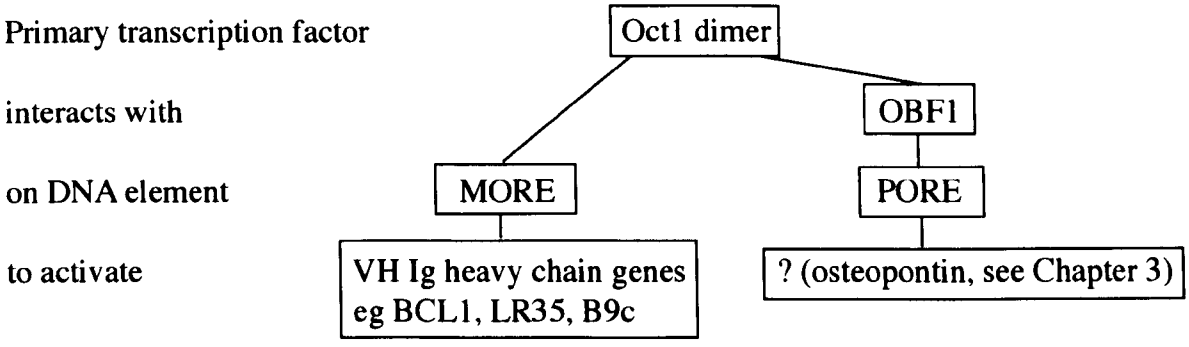
1.2.2 Oct2

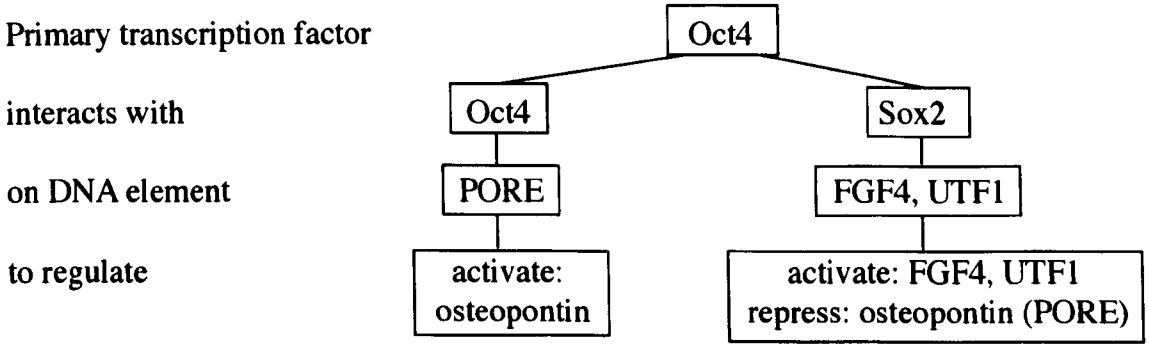
In contrast to Oct1, the expression of Oct2 is restricted to only a few organs and cell types. It was first found in the lymphoid system (B- and T-cells) (Staudt *et al.*, 1986) and later detected in the nervous system, including the developing neural

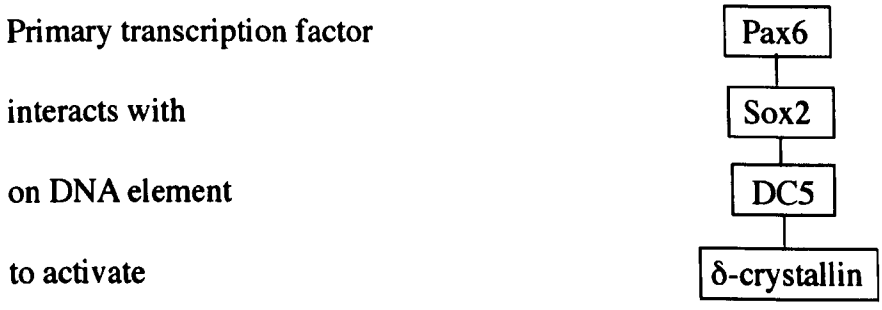
Figure 1.3

Interaction of transcription factors and cofactors

Chart illustrating interaction of different transcription factors and cofactors discussed in this thesis. Their target elements and target genes are also listed. E.g. the Oct1 dimer interacts with OBF1 on the PORE element to activate osteopontin and possibly with an unknown factor on the MORE to activate genes encoding heavy chains of immunoglobulins.


Primary transcription factor
interacts with
on DNA element
to activate
Oct1 dimer
OBF1
MORE
PORE
VH Ig heavy chain genes
eg BCL1, LR35, B9c
? (osteopontin, see Chapter 3)


Primary transcription factor
interacts with
on DNA element
to regulate
Oct4
Oct4
Sox2
PORE
FGF4, UTF1
activate:
osteopontin
activate: FGF4, UTF1
repress: osteopontin (PORE)


Primary transcription factor
interacts with
on DNA element
to activate
Pax6
Sox2
DC5
δ-crystallin

tube and the adult brain, the kidney, the intestine, and testis (Hatzopoulos *et al.*, 1990; He *et al.*, 1989). Oct2 expression and function has been studied in great detail during B-cell differentiation.

The early stages of B-cell differentiation are antigen-independent. They include the three sequential stages: pro-B, pre-B and immature B-cells (Rolink and Melchers, 1991). In the adult mouse these cells are generated in the bone marrow from where the immature B-cells migrate to the lymph nodes and spleen, where they further differentiate and become mature B-cells upon antigen stimulation (Melchers *et al.*, 1995). The differentiation from pro-B to mature B-cells requires DNA rearrangements and specific gene expression profiles. Thus, the various stages of B-cell development are characterized by the rearrangement status of Ig genes as well as the expression of distinct combinations of intracellular or cell-surface markers.

Oct2 is expressed at low levels in pro- and pre-B-cells and at higher levels in more mature B-cells (Staudt *et al.*, 1988; Miller *et al.*, 1991; Corcoran *et al.*, 1993). In pre-B-cells Oct2 expression can be further enhanced by treatment of the cells with e.g. bacterial lipopolysaccharide LPS, a polyclonal B-cell mitogen (Staudt *et al.*, 1986; Staudt *et al.*, 1988).

The function of Oct2 was assessed by gene ablation experiments. Oct2 deficient mice die at birth (Corcoran *et al.*, 1993). The reason is yet unknown but possibly related to a pivotal role of Oct2 in the nervous system (Schubart *et al.*, 2001). Since the focus of Oct2 research lies in the lymphoid system the effect of the ablation there was analyzed in greater detail. It was found that fetal pre-B-cells did not seem to be compromised by the lack of Oct2 (Corcoran *et al.*, 1993). Grafting fetal Oct2 -/- liver to wild-type irradiated mice circumvented the problem of the early death of the animals when analyzing the function of Oct2 in later stages of B-cells in postnatal

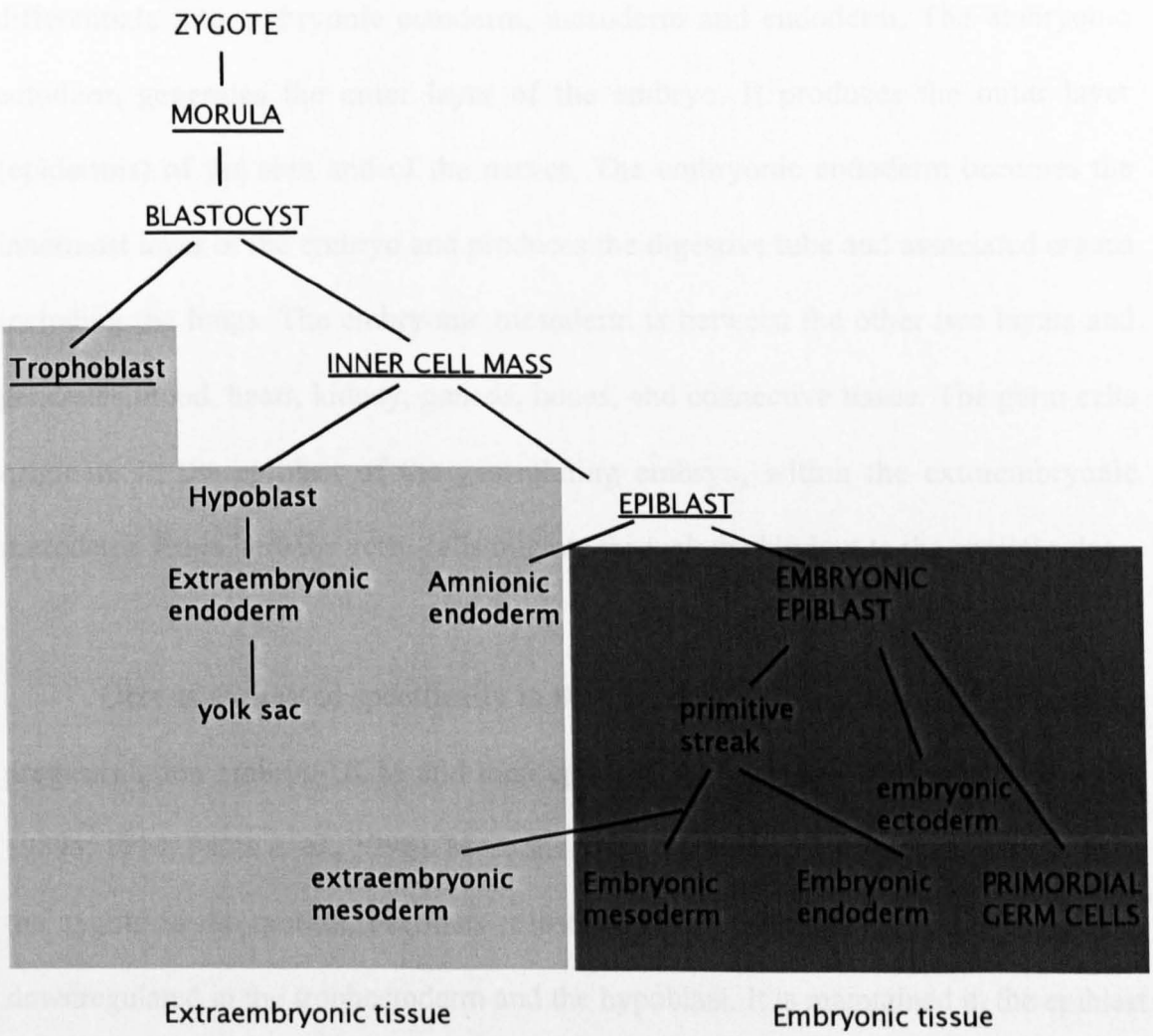
mice (Schubart *et al.*, 2001). The cells exhibit defects in events that commonly occur after B-cell activation, making unique functions for Oct2 in late-stage cells evident. For instance, Oct2 -/- mice have significantly reduced serum Ig levels due to reduced B-cell proliferation upon antigen or LPS activation (Corcoran *et al.*, 1993). This study showed that Oct2 is required for antigen dependent differentiation and proliferation of B-cells. Furthermore, several target genes, have been identified, e.g. CD36, which is activated by Oct2 in B-cells and the tyrosine hydroxylase gene, which is repressed by Oct2 in neuronal cells (König *et al.*, 1995; Dawson *et al.*, 1994). Oct2 also inhibits expression of the herpes simplex virus immediate-early genes, which contain the octamer-related TAATGARAT motif in their promoters (Lillycrop *et al.*, 1991; see Oct1 section 1.2.1).

1.2.3 Oct4

Oct4 is an important factor in early mammalian development. In order to understand its significance I will briefly outline early mammalian development (Gilbert, 2003; figure 1.4). After fertilization the zygote undergoes several successive cleavages. At the 8-cell stage the embryo undergoes compaction where the cells huddle together, maximizing their contact with one another. These cells divide to become the 16-cell morula. The external cells will later become the trophoblast cells, which are not part of the embryo proper, but will give rise to the chorion, the embryonic portion of the placenta. The chorion will enable the connection of the fetus with the uterine wall of the mother. The internal compartment of the compacted morula will become the inner cell mass (ICM), which will give rise to the embryo proper and the yolk sac. During cavitation the trophoblast cells secrete fluid into the morula to create a blastocoel. The resulting structure is called the blastocyst. Now, the ICM is located on one side of the ball of trophoblast cells. The ICM further

Figure 1.4

Schematic diagram showing the derivation of tissues in mammalian embryos (after Gilbert, 2003). Stages at which Oct4 is expressed are written in capital, those at which Sox2 is expressed (see section1.5.2.1) are underlined. (Sox2 is only seen in some cells of the morula, but Oct4 in all cells. Furthermore only the inner cells of the blastocyst express the two transcription factors.)


ZYGOTE
MORULA
BLASTOCYST
Trophoblast
INNER CELL MASS
Hypoblast
EPIBLAST
Extraembryonic endoderm
Amnionic endoderm
EMBRYONIC EPIBLAST
yolk sac
primitive streak
embryonic ectoderm
extraembryonic mesoderm
Embryonic mesoderm
Embryonic endoderm
PRIMORDIAL GERM CELLS
Extraembryonic tissue
Embryonic tissue

differentiates into the hypoblast (primitive endoderm) and epiblast (primitive ectoderm). The hypoblast will contribute to the yolk sac and the epiblast will give rise to the embryo proper and the amnionic ectoderm. The cells of the epiblast that give rise to the embryo proper are called embryonic epiblast. During gastrulation these differentiate into embryonic ectoderm, mesoderm and endoderm. The embryonic ectoderm generates the outer layer of the embryo. It produces the outer layer (epidermis) of the skin and of the nerves. The embryonic endoderm becomes the innermost layer of the embryo and produces the digestive tube and associated organs including the lungs. The embryonic mesoderm is between the other two layers and generates blood, heart, kidney, gonads, bones, and connective tissue. The germ cells originate in the epiblast of the gastrulating embryo, within the extraembryonic mesoderm. From here the germ cells migrate through the hindgut to the genital ridge.

Oct4 is expressed specifically in the germline, i.e. in pluripotent cells of the pregastrulation embryo (ICM and then epiblast) and in germ cells. (Schöler *et al.*, 1989a; 1990; Pesce *et al.*, 1998). More specifically, Oct4 is expressed in all cells from the zygote to the morula, becomes restricted to the cells forming the ICM, and is downregulated in the trophectoderm and the hypoblast. It is maintained in the epiblast and is downregulated at gastrulation in an anterior-posterior pattern excluding the primordial germ cells (PGCs). Subsequently, Oct4 is maintained in the germ cells until the initiation of sexual differentiation of the gametes and meiosis (in the female at 13-14 days post-coitum (dpc) and at the beginning of spermatogenesis in the newborn males). After birth Oct4 is re-expressed in the arrested oocytes. In addition, Oct4 is expressed in cell lines derived from embryos and fetuses, i.e. embryonic stem (ES), embryonic carcinoma (EC) and embryonic germ (EG) cells. As these cell lines

correspond to different stages of the germline, they have turned out to be powerful tools to address molecular biological and biochemical questions (Yeom *et al.*, 1996).

The presence of Oct4 is associated with a pluripotent cell fate (Schöler *et al.*, 1989a; Yeom *et al.*, 1996). Cells in which Oct4 is downregulated differentiate along specific lineages to form embryonic and extraembryonic somatic tissue. Gene ablation experiments have shown that ES cells heterozygous for Oct4 show a 30-40% reduction of this protein and maintain their stem cell characteristics (Nichols *et al.*, 1998). Mice, derived from these ES cells and heterozygous for Oct4 appear normal and are fertile. The absence of Oct4, on the other hand causes early embryonic lethality because cells that should give rise to the ICM of the blastocyst acquire a trophectodermal fate. This indicates that Oct4 is necessary for the maintenance of the pluripotency of embryonic cells. It still has to be shown if Oct4 is required in germ cells.

In contrast to its downregulation during the first differentiation processes, the Oct4 protein level transiently increases during the formation of the hypoblast before it is downregulated (Palmieri *et al.*, 1994; Botquin *et al.*, 1998). This expression profile had been taken as an indication that increased levels of Oct4 may be involved in the establishment of the hypoblast, a notion that has been further supported by experiments in which Oct4 had been inducibly expressed at various levels in ES cells (Niwa *et al.*, 2000). According to this study ES cells require a certain level of Oct4 protein to maintain their cellular identity as ES cells. An increase to more than 150% of normal expression levels in undifferentiated stem cells triggers differentiation into hypoblast and mesoderm, whereas a decrease to less than 50% triggers differentiation into trophectoderm. Overall, Oct4 is assumed to be involved in maintaining pluripotency by two ways of action: by the activation of genes involved in the

maintenance of an undifferentiated state and the repression of genes involved in the differentiation of somatic and extraembryonic cell lineages (Pesce and Schöler 2000).

There are several known target genes of Oct4: FGF4, Osteopontin and UTF1 (figure 1.3) (Yuan *et al.*, 1995; Botquin *et al.*, 1998; Nishimoto *et al.*, 1999).

FGF4 (fibroblast growth factor 4) is coexpressed with Oct4 in the early embryo, ES and EC cells. Later during development it is expressed in the developing limb and tooth but is transcriptionally silent in the adult (Schoorlemmer and Kruijer, 1991; Niswander and Martin, 1992). FGF4 is thought to play a role in embryonic growth prior to gastrulation, induction of the neuroectoderm, neuronal differentiation and survival, and growth and patterning of the developing limb (Goldfarb 1996). FGF4 -/- embryos die at peri-implantation (Feldman *et al.*, 1995). FGFs act by binding to transmembrane cell surface receptors (FGFr), which have intrinsic tyrosine kinase activities.

Osteopontin (OPN), also termed Eta-1 (early T-lymphocyte activation 1) and originally provisionally called 2ar, is a secreted phosphorylated acidic glycoprotein. It is expressed in various tissues, including bone, kidney, cartilage, the placenta, in the inner ear, carcinomas, and is present in most body fluids (Denhardt *et al.*, 1995). Furthermore, OPN is expressed in the immune system, e.g., in T-lymphocytes and activated pro-B-cells (Lin *et al.*, 2000). Secreted OPN stimulates B-cells to produce immunoglobulins and, in conjunction with an unidentified 14kD peptide, to proliferate (Nabel *et al.*, 1981 PNAS; Fresno *et al.*, 1982 Cell). Moreover, B-lymphocytes have been shown to play a role in new bone formation in which OPN is also involved (Marusic *et al.*, 2000). One common theme that has emerged from several of these studies is that OPN is involved in cell migration and adhesion.

UTF1 (**u**ndifferentiated embryonic cell **t**ranscription **f**actor **1**) is a transcriptional cofactor expressed in pluripotential cells, such as those of the inner cell mass (ICM), the epiblast, and in ES and EC cell lines (Okuda *et al.*, 1998). Upon differentiation of these cells UTF1 is no longer expressed. Unlike Oct4 it is later re-expressed in extraembryonic tissues. UTF1 is rich in proline and basic amino acids, especially arginine. It contains two conserved domains, CD1 and CD2 (Fukushima *et al.*, 1998). CD1 is involved in the interaction with TFIID of the basal transcription machinery and is located at the N-terminus of the protein. CD2 contains a leucine zipper motif through which the protein interacts with the metal binding motif of ATF2, which is a leucine zipper transcription factor involved in cellular stress response. UTF1 is thought to act downstream of Oct4 in the regulatory cascade (Okuda *et al.*, 1998).

1.3 HMG proteins

The HMG (high mobility group) superfamily can be divided into two large subfamilies: (i) the classical HMG/UBF family, whose members bind DNA non-specifically; and (ii) the Sox/MATA/TCF family which bind in a sequence specific manner (Laudet *et al.*, 1993). Members of both subfamilies are found in all metazoan species, and play key roles in embryonic development including germ layer formation, cell type specification, and organogenesis. Proteins of the classical HMG subfamily generally contain more than one HMG DNA binding domain and are ubiquitously expressed throughout embryogenesis. Since they bind DNA non-specifically, recognizing unusual DNA structures, such as four-way junctions and kinks, and bend the DNA upon binding, their role in transcriptional regulation was proposed to be primarily architectural (Grosschedl *et al.*, 1994).

Members of the Sox family (reviewed in Wegner, 1999; Pevny and Lovell-Badge, 1997) were originally grouped together on the basis of their high homology to the HMG domain of the testis-determining gene *SRY* (Gubbay *et al.*, 1990; Denny *et al.*, 1992). The similarity to Sry gave them their name Sry box (Sox) proteins. The mammalian Sox family consists of at least 30 genes and can be divided into 10 subgroups based on the homology within and outside of the HMG domain (Bowles *et al.*, 2000).

The HMG domain contains about 80 amino acids and forms an L-shaped module composed of three α-helices (Weir *et al.*, 1993; Chapter 4). Sox factors all recognize similar binding motifs, C(A/T)TTG(A/T)(A/T) to which the HMG domain docks (Harley *et al.*, 1992; Wegner, 1999). All HMG proteins bind on one side of the DNA compressing the major groove and widening the minor groove (Chapter 4; Ferrari *et al.*, 1992; Connor *et al.*, 1994; Love *et al.*, 1995; Werner *et al.*, 1995; Murphy *et al.*, 1999). They induce a sharp bend of the double helix, which can enhance binding of unrelated transcription factors to neighboring DNA sites. The bending angle varies among the different HMG and Sox proteins (Weiss, 2001).

Each Sox factor is expressed in a variety of cellular contexts and a given cell type can co-express a number of Sox factors; their expression is nonetheless restricted temporally and cell specifically.

1.4 Cofactor OBF1 interacts with Oct1 and Oct2

Cofactors facilitate the interaction between the sequence-specific transcription factors and the general RNA polymerase II machinery. OBF1, also termed OcaB and Bob1, is the most extensively studied cofactor of Oct factors (Luo *et al.*, 1992; Gstaiger *et al.*, 1995; Strubin *et al.*, 1995). It has a modular structure that includes a

C-terminal transactivation domain and an N-terminal domain involved in contacting the POU domain (Gstaiger *et al.*, 1996; Cepek *et al.*, 1996). At the molecular level, the exact interaction between OBF1 and the POU domain has been mapped to a short segment in the N-terminal part of OBF1, as directly evidenced by the structure of the monomeric Oct1 in complex with the octamer motif and an N-terminal peptide of OBF1 (figure 1.2B; Chasman *et al.*, 1999). OBF1 also contacts the DNA and requires an adenine at the fifth position of the octamer element (ATGCAAAT).

OBF1 is expressed in lymphoid cells and interacts specifically with Oct1 and Oct2 (Luo *et al.*, 1992; Gstaiger *et al.*, 1995; Luo and Roeder, 1995; Strubin *et al.*, 1995). The interplay of Oct1 and Oct2 with OBF1 is crucial for several aspects of the development and function of B-lymphocytes (Kim *et al.*, 1996; Nielsen *et al.*, 1996; Schubart *et al.*, 1996; Qin *et al.*, 1998; Schubart *et al.*, 2000). In the absence of OBF1, the immune response is dramatically impaired and germinal centers (GCs) do not form. In agreement with this, it was found that expression of OBF1 is upregulated in GC B-cells during the course of an immune response. Germinal centers usually form in the spleen and lymph nodes during the initiation of the acquired immune response (Delves and Roitt, 2000a, 2000b). They create a microenvironment where all necessary antigen-specific and innate antigen-presenting cells can interact. Within GCs hypermutation and class switching of immunoglobulins occur and memory B-cells and plasma-cell precursors are generated. Furthermore, in the absence of OBF1, B-cell development is impaired at the transition between bone marrow and peripheral organs (Kim *et al.*, 1996; Nielsen *et al.*, 1996; Schubart *et al.*, 1996).

Further investigation of Oct1 and OBF1 by Tomilin *et al.* (2000) proposed that the Oct1 or Oct2 dimer is the primary target of OBF1 by showing that the dimer-mediated activation in cotransfection experiments is more enhanced by the cofactor

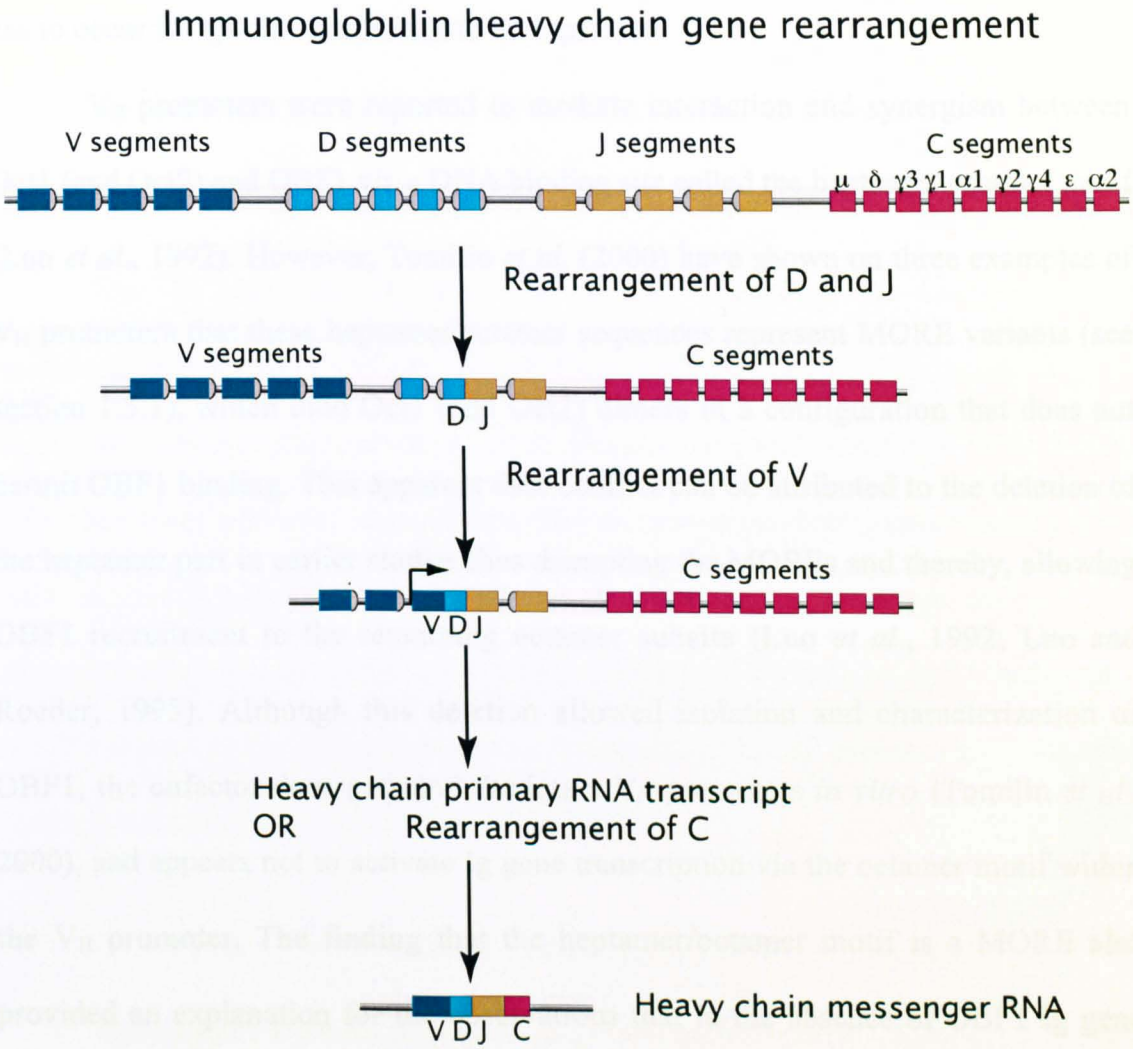
than the monomer-mediated activation. However, the interaction of OBF1 with the Oct1 dimer is dependant on the conformation of the POU/DNA complex, as it can only bind to PORE-mediated dimers and not to those on MOREs. The PORE and the MORE are two different DNA elements on which Oct factors dimerize (see section 1.5.1). The recently solved crystal structures of the POU domains of the Oct1 dimer bound to the MORE and PORE revealed the structural basis for this selectivity of OBF1 (Reményi *et al.*, 2001). The two POU dimers can adopt two different configurations, using different surface patches to dimerize on each element (section 1.5.1). As a result, the same amino acids that are available to interact with OBF1 in the PORE configuration (Chasman *et al.*, 1999) form part of the POU_S-POU_H dimer interface on the MORE, precluding an interaction with OBF1 (figures 1.1 and 1.2B) (Tomilin *et al.*, 2000, Reményi *et al.*, 2001).

Another reason for the impaired immune response, besides the lack of germinal center formation, in OBF1-/- mice is the deficiency of immunoglobulin production commonly known as antibodies. Immunoglobulins (Ig) are glycoproteins consisting of 2 identical heavy chains and 2 identical light chains (Delves and Roitt, 2000a). The light chains consist of 3 regions (V: variable, D: diversity and J: joining) and are divided into 2 types, κ and λ. Recently, OBF1 was shown to be essential for transcription and VDJ recombination of a subset of Igκ genes *in vivo* (Casellas *et al.*, 2002).

The heavy chains (IgH) consist of 4 regions (V, D and J like the light chains and a constant region C) and are divided into 5 classes (IgG, IgA, IgM, IgD and IgE) depending on which constant region was recombined to the VDJ (figure 1.5; Arulampalam *et al.*, 1997). All heavy chain mRNA transcripts are regulated by a

Figure 1.5

Heavy and light chain genes recombine in pre-B-cells (after Delves and Roitt, 2000a). The recombination enables an efficient immune response upon infection of the very diverse pathogens. After recombination a specific mature B-cell only expresses one type of antibody. Clonal expansion of this B-cell upon infection ensures that the pathogen can be eliminated as soon as possible.


Immunoglobulin heavy chain gene rearrangement
V segments
D segments
J segments
C segments
μ δ γ3 γ1 α1 γ2 γ4 ε α2
Rearrangement of D and J
V segments
C segments
D J
Rearrangement of V
C segments
V D J
Heavy chain primary RNA transcript
OR
Rearrangement of C
V D J C
Heavy chain messenger RNA

heavy chain (V_H) promoter, which is located 5' of the V segment. Every V segment has its own specific V_H promoter. The V_H promoters become fully active when recombination brings them into proximity of an enhancer in the C gene segment. The V_H promoter closest to the C enhancer is then active after VDJ recombination, which has to occur for specific antibodies to be expressed.

V_H promoters were reported to mediate interaction and synergism between Oct1 (and Oct2) and OBF1 via a DNA binding site called the heptamer-octamer motif (Luo *et al.*, 1992). However, Tomilin *et al.* (2000) have shown on three examples of V_H promoters that these heptamer/octamer sequences represent MORE variants (see section 1.5.1), which bind Oct1 (and Oct2) dimers in a configuration that does not permit OBF1 binding. This apparent data conflict can be attributed to the deletion of the heptamer part in earlier studies thus disrupting the MOREs and thereby, allowing OBF1 recruitment to the remaining octamer subsite (Luo *et al.*, 1992; Luo and Roeder, 1995). Although this deletion allowed isolation and characterization of OBF1, the cofactor does not bind the intact V_H promoters *in vitro* (Tomilin *et al.*, 2000), and appears not to activate Ig gene transcription via the octamer motif within the V_H promoter. The finding that the heptamer/octamer motif is a MORE also provided an explanation for the observations that in the absence of OBF1 Ig gene transcription is only marginally impaired (Kim *et al.*, 1996; Nielsen *et al.*, 1996; Schubart *et al.*, 1996; Schubart *et al.*, 2001). The gene ablation experiments did show though that OBF1 exerts an effect on Ig heavy chain transcription after VDJ recombination via a 3' distal enhancer that contains sequences OBF1 can get recruited to. The Ig locus rearrangement activates the 3' enhancer by bringing it into the proximity to the V_H promoters. On another note, Oct1 stimulates the MORE-containing V_H promoters less effectively in non-lymphoid than in B-cells,

suggestingthe existence of a lymphoid-specific cofactor other than OBF1, as proposed in earlier work (Kim *et al.*, 1996; Schubart *et al.*, 1996; Schubart *et al.*, 2001). This hypothetic cofactor should interact with Oct1 bound in the MORE configuration and play a role in Ig gene transcription.

1.5 Factors affecting transcriptional activation of POU protein activity

The 15 members of the POU family identified in mouse have a multifunctional capacity and are critical regulatory components of a broad range of biological processes (see section 1.2.1-3). Despite the small number of POU factors their regulatory potential is enormous due to inter-dependant control mechanisms such as postranslational modifications, flexible DNA sequence recognition, oligomerization, and interaction with transcriptional cofactors (Herr and Cleary, 1995 and this thesis). Many of these properties are conferred by the evolutionary conserved bipartite DNA binding structure, the POU domain.

1.5.1 Dimerization: Two DNA motif-induced protein conformations

Oct proteins were initially identified to act as monomeric transcription factors, binding to the octamer sequence and related DNA motifs. Dimerization on certain DNA elements such as the heptamer/octamer motif have also been described, but structure-functional aspects have not been characterized to the same extent (reviewed in Herr and Cleary, 1995). More recently, DNA-mediated dimerization of POU factors and its impact on eukaryotic gene expression received substantial attention. (Jacobson *et al.*, 1997; Botquin *et al.*, 1998; Rhee *et al.*, 1998; Scully *et al.*, 2000; Tomilin *et al.*, 2000; Reményi *et al.*, 2001) Pit1 homodimers and POU heterodimers

with other transcription factors such as Sox2 had been identified earlier (Ingraham *et al.*, 1990; Dailey *et al.*, 1994).

For Oct factors two types of DNA elements, the PORE (**P**alindromic **O**ct factor **R**ecognition **E**lement, ATTTGAAA[T/G]GCAAAT, 15bp) and the MORE (**M**ore **PORE**, AT[G/A][C/A]AT[N]$_{0-3}$ATGCA[A/T], 12-15 bp) have been characterized (Botquin *et al.*, 1998; Tomilin *et al.*, 2000; Reményi *et al.*, 2001).

The PORE was first identified as an Oct4 binding sequence in the first intron of the *opn* gene (Botquin *et al.*, 1998). Homo- and heterodimers of various POU factors, including Oct1, Oct2 and Oct4 can assemble on the PORE *in vitro*, and cotransfection experiments using a series of PORE mutations suggest that dimer formation also occurs in cells. However, it has not yet been proven if POU dimers indeed form *in vivo* and it is far from clear if POU homo- or heterodimers play a role during mammalian development. MOREs and related sequences (e.g., the Heptamer/Octamer motif) are found in immunoglobulin heavy chain promoters (see section 1.4; Tomilin *et al.*, 2000). Oct family members can also bind cooperatively as homo- and heterodimers on the MORE *in vitro*.

The quaternary arrangements of the four POU subdomains in a dimer, two POU_S and two POU_H, are different for the MORE and PORE type dimers. This was first shown by a series of biochemical experiments and ultimately determined by solution of X-ray structures of the dimeric Oct1/PORE and Oct1/MORE complexes (figure 1.2C; Reményi *et al.*, 2001). The POU_S and POU_H subdomains contain two different non-overlapping pairs of surface patches that are capable of forming unrelated dimerization interfaces, dictated by the DNA sequence.

In the PORE configuration, the two main protein-protein interfaces formed in the complex are located across the semi-palindromic center of the PORE motif, thus

each protein extends across the longitudinal axis of the DNA molecule (see figure 1.2C; Botquin *et al.*, 1998; Reményi *et al.*, 2001). The N-terminus of POU_H of one molecule interacts with residues of α-helices 1 and 2 of POU_S of the other molecule. In the MORE configuration, on the other hand, the interface between the POU_S and POU_H domains is located within each half-site of the DNA element therefore each protein molecule is arranged along the longitudinal axis of the DNA (see figure 1.2C). The C-terminal residues (157-160) of POU_H of one molecule dock onto a loop region between α-helices 3 and 4 of POU_S of the other molecule. The Oct1 side chain of I159 (S159 in Oct4) fits snuggly into a hydrophobic pocket consisting of residues of POU_S. With these findings Reményi *et al.* (2001) introduced the concept of interface swapping in dimers as a general mechanism of modulating the activity of transcription factors.

1.5.2 Heterodimerization with Sox proteins

Oct factors may heterodimerize with transcription factors not belonging to the POU family. In order for Sox proteins to exert their specificity, it has been proposed previously that they combine with different protein partners to direct gene expression required for a particular cell type. Indeed, interaction and/or cooperation between Sox factors and various POU factors have been observed, including Sox10 with Tst1/Oct6/SCIP, Sox11 and Sox4 with Brn1, and Sox2 with Oct4 (Yuan *et al.*, 1995; Botquin *et al.*, 1998; Nishimoto *et al.*, 1999; Kuhlbrodt *et al.*, 1998a,b).

1.5.2.1 Sox2

Sox2 is coexpressed with Oct4 in the ICM, the epiblast and germ cells. Consistent with this, Sox2 and Oct4 are also coexpressed in ES and EC cells (Botquin *et al.*, 1998; Chapter 5).

Unlike Oct4, Sox2 is expressed in multipotent trophoblast cells and the chorion. Sox2 is downregulated in both the embryonic and the extraembryonic lineages as the cells differentiate and lose their multipotency (Avilion *et al.*, 2003). Later, Sox2 is expressed in the developing nervous system (Collignon *et al.*, 1996) and the lens (Kamachi *et al.*, 1995).

The expression of genes such as fibroblast growth factor 4 (*FGF4*) and undifferentiated embryonic cell transcription factor 1 (*UTF1*), which are coexpressed with Sox2 and Oct4, depends on enhancer elements that contain binding sites for these HMG and POU transcription factors (Yuan *et al.*, 1995; Nishimoto *et al.*, 1999; section 1.2.3). Sox2 not only acts as a transcriptional activator but it can also repress transcription in some contexts, as observed in the regulation of the *OPN* gene early in development (figure 1.3) (Botquin *et al.*, 1998).

Sox2 homozygous null embryos die soon after implantation, revealing that Sox2 is essential during embryogenesis (Avilion *et al.*, 2003). Despite its importance in early developmental gene regulation, little is known about the transcriptional regulation of the *Sox2* gene itself. Only recently Sox2 was found to be regulated by the Oct4/Sox2 heterodimer in undifferentiated ES and EC cells (Tomioka *et al.*, 2002).

1.5.3 Extension of the network: Sox proteins interact with Pax proteins

Members of the Sox protein family also interact with members of the Pax protein family. For example, Sox2 interaction with Pax6 (see below) is crucial for proper eye development in mammals. Pax proteins possess a 128 amino acid DNA binding domain, the 'paired domain' (PD) and a C-terminal transactivation domain (Stuart *et al.*, 1994). Pax proteins are known to play critical roles in mammalian development and oncogenesis (Chi and Epstein, 2002).

The PD consists of two subdomains which both structurally resemble a helix–turn–helix (HTH) motif, similar to the HTH motif found in homeodomains (Xu *et al.*, 1995, 1999). The subdomains are connected by a flexible linker. Several Pax proteins, including Pax6, also contain an additional homeodomain (HD). Cooperative dimerization with the paired domain of the same molecule allows these homeodomains to recognize sequences not bound by other HDs (Jun and Desplan, 1996; Underhill and Gros, 1997).

Pax6 is expressed in the nose, pancreas, and coexpressed with Sox2 in the developing eye and the central nervous system (Walther and Gruss, 1991; Turque *et al.*, 1994; Kamachi *et al.*, 1995; Collignon *et al.*, 1996; Grindley *et al.*, 1997). Pax6 and Sox2 bind cooperatively to the DC5 DNA element to synergistically activate the expression of the δ-crystallin gene (Kamachi *et al.*, 2001; figure 1.3), which is essential for lens development during late embryogenesis. Formation of the ternary complex involves direct protein–protein interactions between the Pax6 paired domain and the HMG domain of Sox2 (Chapter 4). Exogenous Sox2 and Pax6 in the head ectoderm can induce ectopic lens placode development showing the significance of Pax6 and Sox2 for initiation of lens development.

The additional identification of Pax3 as an interactive partner of Sox10 suggests that various Pax and Sox factors might work in combination to affect tissue-specific gene expression (Bondurand *et al.*, 2000; Lang and Epstein 2003).

1.5.4 Phosphorylation

One of the most frequently used ways to regulate the activity of transcription factors in response to different extra- and intracellular signals is phosphorylation and

dephosphorylation (Whitmarsh and Davis, 2000). This is also true for the POU family of transcription factors. For example, several members become hyperphosphorylated during the M-phase of the cell cycle (Roberts *et al.*, 1991; Segil *et al.*, 1991; Caelles *et al., 1995).* The purpose of this mitotic phosphorylation is likely to be the prevention of the POU protein from binding to DNA, thereby precluding transcription as the cells enter mitosis. Entry into the G1 growth phase is accompanied by dephosphorylation and regain of DNA binding activity and transcription. One target for phosphorylation was identified within the POU homeodomain at serine 107 (Oct1) or threonine 107 (Pit1), which is located very close to the DNA when the protein is bound (figure 1.1 and 1.2A). Phosphorylation of this residue leads to reduced DNA binding (Kapiloff *et al.*, 1991; Segil *et al.*, 1991; Caelles *et al.*, 1995), probably because the negative charge in the phosphate is repulsed by the negative DNA backbone.

Besides cell-cycle dependent phosphorylation of POU proteins, which seems to disrupt overall DNA binding, several observations suggested that specific target genes are regulated by POU phosphorylation upon cell signals. Pit1 binding sites have been implicated in regulating prolactin and growth hormone promoter activity in response to agents activating signal transduction pathways such as cAMP, phorbol esters (TPA) and activin A. The latter three are known to activate specific protein kinases. Treatment of cells with either of these signaling agents activates specific cellular protein kinases leading to increased Pit1 phosphorylation. This in turn was found to affect Pit1 binding to particular sites (Kapiloff *et al.*, 1991; Struthers *et al.*, 1992; Caelles *et al.*, 1995). Kapiloff *et al.* (1991) particularly showed how phosphorylation differentially affects protein binding to DNA response elements depending on their nucleotide sequence. Thus, signal-induced phosphorylation of transcription factors may be a means of differentially regulating transcription of two

genes, which are under the influence of the same transcription factor. *In vitro,* Pit1 can be phosphorylated by cAMP dependent protein kinase PKA and PKC. Analysis of the PKA phosphorylation sites showed that one site was identical to the mitotic phosphorylation site T107 (Kapiloff *et al.*, 1991; Segil *et al.*, 1991; Roberts *et al.*, 1991). PKA, however, is not accountable for the M-phase phosphorylation because specific inhibition of PKA did not block Pit1 phosphorylation by mitotic extracts (Caelles *et al.*, 1995). The *in vivo* relevance of PKA in mediating Pit1 activation still needs to be proven.

1.6 Aim of this thesis

It has been established that transcription factors often interact with each other but the exact modes of interaction and target genes mostly still remain to be investigated.

The overall objective of this thesis is to elucidate how Oct transcription factors interact with other transcription factors or cofactors to regulate transcription. I chose two paradigmatic POU factors, Oct1 and Oct4, to discover novel target transcriptional functions, and assess how their activities are influenced by multi-level control mechanisms. More specifically the scope of the work was to investigate the interaction of Oct1 with the B-cell specific cofactor OBF1 and the interaction of Oct4 with Sox2.

Based on the knowledge that OBF1 specifically interacts with PORE-type Oct1 dimers but not with those on the MORE (Tomilin *et al.*, 2000; Reményi *et al.*, 2001) the effect of OBF1 binding on the PORE-type Oct1 dimer was to be investigated further. To this end the effect of OBF1 on Oct1 dimer interface mutants

and DNA sequence requirements had to be studied. Furthermore, the biological relevance of this Oct1-OBF1 interaction was sought (**Chapter3**).

Another focus was to study the way of Oct4 and Sox2 regulating various early embryonic stage specific enhancers:

The following question arose from the knowledge that Oct4 and Sox2 activate FGF4 and UTF1 expression (Yuan *et al.*, 1995; Nishimoto *et al.*, 1999): how do Sox2 and Oct4 interact on these DNA binding elements and is the interaction similar or identical for the two enhancers? The solution of a crystal structure and the use of models would give a basis for biochemical experiments to address these issues (**Chapter 4**).

Whereas Oct4 and Sox2 activate FGF4 and UTF1 expression, Sox2 represses Oct4 activation of osteopontin (Botquin *et al.*, 1998). This calls to examine the mechanism of repression. Attending to this issue, the physical interference of Sox2 with Oct4 mediated activity should be solved (**Chapter 5**).

Chapter 2

Material and Methods

2.1 Materials

2.1.1 Chemicals

Powder for bacterial media and agar was purchased from Fisher Scientific, USA (LB agar: BP-1425-2; LB broth: BP-1426-2) Cell culture media were obtained from Gibco-BRL and Sigma. All other organic and inorganic chemicals were obtained from Sigma unless specified otherwise. All protocols use ultrapure water (Millipore).

2.1.2 Radioactive isotopes

(α-^{32}P)dCTP (10mCi/ml) was obtained from Amersham Bioscience, USA (# AA0075-250μCi) for labeling double-stranded (ds) oligonucleotides (section 2.3.8) for electrophoretic mobility shift assays (section 2.8).

2.1.3 Enzymes

Restriction enzymes, DNA polymerase I Klenow fragment, calf intestinal phosphatase (CIP), T4 DNA ligase, and T4 DNA polymerase were purchased from New England Biolabs. *PfuTurbo*® DNA polymerase was obtained from Stratagene.

2.1.4 Synthetic oligonucleotides

Oligonucleotides were synthesized by Invitrogen™ Life Technologies and desalted.

For the DNA-binding studies using double-stranded oligonucleotides, the complementary strands were synthesized such that upon hybridization a 4 base (CTGA) 5' overhang at each end occurred. The top strands of the double stranded oligonucleotides are indicated below. The PORE and MORE elements, Sox and Pax binding sites are written in capital letters, the octamer motif is underlined, the Sox

binding site is double underlined and mutations are written bold. Some of these sequences are already published in Botquin *et al.* (1998), Tomilin *et al.* (2000), Kamachi *et al.* (2001).

Oligonucleotides used in mobility shift assays:

Igκ: 5'-ctgactcctgccttcagggtATGCAAATtattaagtctcgag-3'

PORE, DM: 5'-ctgaaagttaaaatcacATTTGAAATGCAAATggaaaagcaag-3'

PORE^{D}, Dm: 5'-ctgaaagttaaaatcacATTTGAAAGGCAAATggaaaagcaag-3'

PORE^{M}, dM: 5'-ctgaaagttaaaatcacATGTGAAATGCAAATggaaaagcaag-3'

PORE-A1C: 5'-ctgaaagttaaaatcacATTTGAACTGCAAATggaaaagcaag-3'

PORE-A5T: 5'-ctgaaagttaaaatcacATTTGAAATGCTAATggaaaagcaag-3'

PORE-A6G: 5'-ctgaaagttaaaatcacATTTGAAATGCAGATggaaaagcaag-3'

PORE-A7C: 5'-ctgaaagttaaaatcacATTTGAAATGCAACTggaaaagcaag-3'

P+1: 5'-ctgaaagttaaaatcacATTTGATAATGCAAATggaaaagcaag-3'

P-1: 5'-ctgaaagttaaaatcacATTTGA ATGCAAATggaaaagcaag-3'

PORE^{D}->Igκ: 5'-ctgaaagttaaaatcacATTTGAAAGGCAAATtaaaaagcaag-3'

5SDM 5'-ctgaaaagatatCTTTGTTtctttATTTGAAATGCAAATggaaaagc-3'

MORE: 5'-ctgaaagttaaaatctcATGCATATGCATggaaaagcaag-3'

BCL1: 5'ctgacctgtctcATGAATATGCAAatcaggtgagtctatggtggtaaatatagggatatca-3'

FGF4: 5'-ctgaaagaaaactCTTTGTTtggATGCTAATgggatactaagctga-3'

UTF1: 5'-ctgaaagatgagagccctCATTGTTATGCTAGTgaagtgccaagctga-3'

DC5: 5'-ctgatattCATTGTTGTTGCTCACCTACCATGGATCCGAActga-3'

1W: 5'-ctgacttattttagaaATGCAAATtacccaggtggtgtttgc-3'

Oligonucleotides used for Transformer™ site-directed mutagenesis (section 2.3.6) have a 5' phosphate group attached. Those used for QuikChange® site directed

mutagenesis (section 2.3.7) do not have the 5' phosphate group. Furthermore, only the reverse primer (R) plus the AlSpSDMtogl primer were used for the Transformer™ site-directed mutagenesis, whereas the upper and lower primer were used for the QuikChange® site directed mutagenesis. Mutations from the original sequences are indicated by bold letters; the underlined sequences are the codons encoding an amino acid supposedly involved in a protein-protein interface mutated to another (indicated in nomenclature); the double underlined sequence represents the introduction of a restriction site (AlwNI instead of SpeI).

AlwNI-SpeI toggle:

5'p G**CA GCC ACT GGT A**AC AGG ATT 3'

POU1m1 (Q18A, I21A)

F: 5'-GTTTGCCAAGACCTTCAAA**GCA**AGACGA**GCC**AAACTTGGATTCACTCAG-3'

R: 5'-CTGAGTGAATCCAAGTTTGGCTCGTCTTGCTTTGAAGGTCTTGGCAAAC-3'

POU1m3 (I21Y)

F: 5'-AGCCTTCAAACAAAGACGA**TAC**AAACTTGGATTCACTCAG-3'

R: 5'-CTGAGTGAATCCAAGTTTGTATCGTCTTTGTTTGAAGGCT-3'

POU1m4 (I21G, K22A)

F: 5'-CCTTCAAACAAAGACGA**GGCGCA**CTTGGATTCACTCAGG-3'

R: 5'-CCTGAGTGAATCCAAGTGCGCCTCGTCTTTGTTTGAAGG-3'

POU1m1 (K104S, E109A)

F: 5'-GAGCCGTAGGAGGAAG**TCC**CGCACCAGCATA**GCG**ACCAACATCCGTGTGGC-3'

R: 5'-GCCACACGGATGTTGGTCGCTATGCTGGTGCGGGACTTCCTCCTACGGCTC-3'

POU1m11 (S107A)

F: 5'-GGAGGAAGAAACGCACC**GCC**ATAGAGACCAACATCCG-3'

R: 5'-CGGATGTTGGTCTCTATGGCGGTGCGTTTCTTCCTCC-3'

POU1m12 (S107D)

F: 5'-GGAGGAAGAAACGCACC **GAC** ATAGAGACCAACATCCG-3'

R: 5'-CGGATGTTGGTCTCTATGTCGGTGCGTTTCTTCCTCC-3'

POU1m13 (S107E)

F: 5'-GGAGGAAGAAACGCACC **GAA** ATAGAGACCAACATCCG-3'

R: 5'-CGGATGTTGGTCTCTATTTCGGTGCGTTTCTTCCTCC-3'

Sox2 R75E:

F: 5'-GAT TAT AAA TAC CGG CCG **GAG** CGC AAA ACC AAG-3'

R: 5'-CTT GGT TTT GCG CTC CGG CCG GTA TTT ATA ATC-3'

Sox2 K57E, R60E:

F: 5'-CCGTTCATCGACGAGGCC**GAG**CGGCTG**GAA**GCTCTGCACATGAAG-3'

R: 5'-CTTCATGTGCAGAGCTTCCAGCCGCTCGGCCTCGTCGATGAACGG-3'

Sox2 R60E, M64E:

F: 5'-GAG GCC AAG CGG CTG **GAA** GCT CTG CAC **GAG** AAG GAG CAC CCG GAT TAT AAA TAC-3'

R: 5'-GTA TTT ATA ATC CGG GTG CTC CTT CTC GTG CAG AGC TTC CAG CCG CTT GGC CTC-3'

POU1 I21Y, D29R:

F: 5'-CC TTC AAA CAA AGA CGA **TAC** AAA CTT GGA TTC ACT CAG GGT **CGT** GTT GGG CTC GCT ATG GGG-3'

R: 5'-CCC CAT AGC GAG CCC AAC ACG ACC CTG AGT GAA TCC AAG TTT GTA TCG TCT TTG TTT GAA GG-3'

POU4/Oct4 I21Y, D29R:

F: 5'-CTG CTG AAG CAG AAG AGG **TAC** ACC TTG GGG TAC ACC CAG GCC **CGT** GTG GGG CTC ACC CTG-3'

R: 5'-CAG GGT GAG CCC CAC ACG GGC CTG GGT GTA CCC CAA GGT GTA CCT CTT CTG CTT CAG CAG-3'

Oligonucleotides used for annealing with complementary oligonucleotides (giving rise to 5' CTGA-overhangs) and subsequent hexamerization of part of *osteopontin*'s intron 1 are shown below. Hexamers were then cloned into *tk* luciferase reporter plasmids. Sox sites and POREs are underlined, mutations are indicated by bold letters, lowercase letters are non-*opn* sequences.

SDM: 5'-ctgaTCTTTGTTTCTTTCAGCTTTGTATAATGTAAGTTAAAATCACATTTGAAATGCAAATGGAAAAGCaagtcga-3'

sDM: 5'-ctgaT**GCACTGAC**CTTTCAGCTTTGTATAATGTAAGTTAAAATCACATTTGAAATGCAAATGGAAAAGCaagtcga-3'

SDm: 5'-ctgaTCTTTGTTTCTTTCAGCTTTGTATAATGTAAGTTAAAATCACATTTGAAA**GG**CAAATGGAAAAGCaagtcga-3'

SdM: 5'-ctgaTCTTTGTTTCTTTCAGCTTTGTATAATGTAAGTTAAAATCACATGTGAAATGCAAATGGAAAAGCaagtcga-3'

sDm: 5'-ctgaT**GCACTGAC**CTTTCAGCTTTGTATAATGTAAGTTAAAATCACATTTGAAA**GG**CAAATGGAAAAGCaagtcga-3'

sdM: 5'-ctgaT**GCACTGAC**CTTTCAGCTTTGTATAATGTAAGTTAAAATCACATGTGAAATGCAAATGGAAAAGCaagtcga-3'

5SDM: 5'-ctgaAAAGATATCTTTGTTTCTTTATTTGAAATGCAAATGGAAAAG Cagtcga-3'

11SDM: 5'-ctgaAAAGATATCTTTGTTTGTTAAATCACATTTGAAATGCAAAT GGAAAAGCagtcga-3'

2.1.5 Synthetic oligopeptide

The OBF1 peptide used in Chapter 3 (MLWQKPTAPEQAPAPARPYQGVR VKEPVKELLRRKRGHASSSSGAA) was chemically synthesized by Sigma-Genosys, USA, and is identical to the peptide used for crystallization by Chasman *et al.* (1999). Its composition and purity were checked by HPLC and Mass Spectral analysis. The lyophilized peptide was resuspended in 10mM HEPES (pH7.6), 100mM NaCl, 5mM β-mercaptoethanol (β-ME) and stored at -80°C.

2.1.6 Antibodies

Immunization of rabbits with bacterially expressed Oct4 protein and preparation of antibody was performed by Karin Hübner at the EMBL, Heidelberg. Sox2 antibody was kindly provided by Lisa Dailey, NYU, New York. Oct1 antibody was purchased from Santa Cruz, USA (C-21: sc-232). Antibodies were used to supershift protein-DNA complexes in EMSAs (section 2.8) and to detect the corresponding proteins by Western blot analyses (section 2.6.3).

2.1.7 Other materials

Kodak X-OMAT™ Blue XB-1 and Biomax™ ML films were purchased from Sigma USA. Dialysis tubing and membranes were purchased from Pierce, USA. Nylon membrane Hybond-N+ and ECL™ Western Blotting Detection Reagents were

purchased from Amersham Pharmacia. QIAGEN Plasmid Kits and QIAquick™ Gel Extraction Kits (Cat# 28704) were purchased from Qiagen, Germany.

2.2 Bacterial techniques

2.2.1 Bacterial strains

The *E.coli* XL 1-Blue strain (genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q}ZΔ*M15* Tn*10* (Tetr)]) was used for propagation of double stranded plasmids.

At other times the *E.coli* DH5α strain (genotype: *sup E44ΔlacU169 (ϕ80lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*) was used for propagation of double stranded plasmids.

The mismatch repair deficient *E.coli* BMH 71-18 *mut*S strain (genotype: *thi, sudE,* Δ(*lac-proAB*), [*mut*S::Tn10][F'*proAB, lac*I^{q}ZΔM15] was used after site-directed mutagenesis with the Clontech kit.

The BL21-CodonPlus(DE3)-RIL strain (Stratagene #230345) (genotype: *E. coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal* λ (DE3) *endA* Hte *[argU ileY leuW* Camr]) was used for bacterial protein expression. It is a general protein expression strain that lacks both the Lon protease and the OmpT protease, which can degrade recombinant proteins during expression and purification. The naturally lacking Dcm methylase is inserted into the genome. Efficient expression of heterologous proteins is often limited by the rarity of certain tRNAs in the bacteria that are abundant in the organisms from which the proteins are derived. BL21-CodonPlus-RIL cells contain extra copies of the *argU, ileY,* and *leuW* tRNA genes, which recognize the AGA/AGG, AUA, and CUA codons, respectively, facilitating the expression of heterologous genes, which contain these rare codons. The DE3 lysogen contains the

T7 polymerase gene, which is under the control of the IPTG inducible *lacUV5* promoter. The polymerase in turn activates protein expression driven by T7 promoters, such as in pET vectors used in this thesis.

2.2.2 Maintenance and Media

E.coli cells were grown at 37°C (unless specifically mentioned otherwise) in L-broth, shaking at 200 rpm, or on LB agar plates containing antibiotics when needed as described in Sambrook *et al.* (1989). Carbenicillin was used instead of ampicillin.

2.2.3 Preparation of competent *E.coli* cells

The protocol for competent cells (chemical transformation) published in the user manual of the Transformer™ Site-Directed Mutagenesis Kit from Clontech (Catalog #K1600-1) was used for XL 1-Blue, DH5α, BMH 71-18 *mut*S and BL21-CodonPlus(DE3)-RIL cells. Only the BMH 71-18 *mut*S bacteria were grown on tetracycline LB plates and in tetracycline L-broth.

2.2.4 Chemical transformation of competent *E.coli* cells

A 100μl aliquot of competent cells was thawed on ice. Then 5-10μl DNA was added to the bacteria in the Eppendorf tube, mixed gently and incubated on ice for 30'. After a 1' heat shock at 42°C, cells were allowed to recover with 400μl LB media at 37°C for 1hr. Typically 100μl of the culture were plated onto an LB plate containing the appropriate antibiotic. The plates were incubated overnight (12-16hrs) at 37°C.

2.3 Recombinant DNA Techniques

2.3.1 Phenol/Chloroform extraction and ethanol precipitation of DNA

This procedure was carried out routinely to remove proteins from DNA containing solutions. The volume of the solution was brought to 200-500µl with ultrapure H_2O, proteins were then removed with the help of phenol/chloroform as described by Sambrook *et al.* (1989). The DNA was precipitated by adding 0.1 volumes 3M Na-acetate (pH5.2) and 2.5 volumes ice-cold ethanol. The solution was mixed and incubated at -20°C for 5'. The DNA was recovered by centrifugation in a cooled microcentrifuge (15'at 13,000rpm) The DNA pellet was washed with 500µl 70% ethanol, air-dried and resuspended in an appropriate volume of TE (pH8.0; 10mM Tris pH8.0, 1mM EDTA).

2.3.2 CsCl purification of DNA

CsCl purification of DNA started by spinning 300ml bacteria culture (ThermoForma centrifuge model 5696, swinging bucket rotor (No. 5808947), 3000rpm) for 15' at 4°C. The pellet was resuspended in 5ml lysozyme compatible buffer (50mM glucose, 10mM EDTA, 25mM Tris (pH 8.0)) and transferred to a 50ml Falcon tube. Then another 5ml lysozyme compatible buffer containing 100mg lysozyme was mixed in well and incubated on ice for 5'. Next, 10ml 0.2M NaOH, 1% SDS were added, mixed immediately by gentle shaking and incubated on ice for 5-10'. Then 8.4ml KOAc (3M with respect to potassium, 5M with respect to acetate; Sambrook *et al.*, 1989) were mixed in and incubated on ice for 15'. Then the solution was centrifuged for 15' at 4°C (see conditions above). The supernatant was transferred to a new 50ml Falcon tube. If necessary, autoclaved gauze was used to remove floating debris. Next, 0.6 volumes isopropanol were added and mixed in to

precipitate the DNA by incubating the mixture on ice for 30'. Then the solution was centrifuged again as above. The pellet was air-dried moderately and resuspended in 3ml TE (pH8.0). RNase A was added to a final concentration of 125µg/ml and incubated for 60' at RT. For each ml of TE 1g CsCl and 50µl EtBr (10mg/ml) were added. The samples were spun for 10' at RT. Using a Pasteur pipet the supernatants were transferred to polyallomer centrifugation tubes (Beckman Coulter Cat. # 326819), which were then balanced. The samples were centrifuged in a Beckman Coulter™ Optima™ MAX-E ultracentrifuge (swinging bucket rotor MLS-50) for 20hrs at 40,000 rpm. The supercoiled plasmid was recovered by puncturing a syringe with an 18 gauge needle directly underneath the lower pink band and removing it (see Sambrook *et al.* (1989). The EtBr was removed by extracting the DNA solution 3 times with 3 volumes of n-BuOH saturated with TE. After all visible EtBr was extracted the DNA solution was dialyzed against TE at RT with 3 changes (1L, 1hr, each). Then the DNA was precipitated with 0.1 volumes 3M Na-acetate (pH5.2) and 2.5 volumes ethanol. The DNA pellet was washed with 500µl 70% ethanol, air-dried and resuspended in 100µl TE (pH8.0; 10mM Tris pH8.0, 1mM EDTA).

2.3.3 Cloning strategy

All DNA used for cloning was dissolved in TE. Precautions were taken to maximize the generation of recombinant plasmids: Both vector and insert fragments were purified by gel extraction with the QIAquick™ Gel Extraction kit (Cat.# 28704) to exclude the presence of undigested plasmids in the ligation reaction. Recircularization of vector fragments was minimized in two ways: (a) Where possible DNA fragments were inserted into vectors cut with two enzymes generating incompatible ends. (b) The 5'phosphate moieties of the vector fragments were removed with calf intestinal phosphatase (CIP). As ligation by T4 DNA ligase

requires the presence of 5'phosphate groups on only one of the two DNA fragments this treatment counteracts recircularization of the vector without significantly affecting ligation of untreated insert with treated vector.

(i) Preparation of vector and insert: 3-10μg vector plasmid and insert carrying plasmid were digested with 20-30 units of the appropriate restriction enzymes for 1-2hrs. Digestions with two restriction enzymes requiring different digest buffers were carried out sequentially. If compatible ends were present, 1 unit CIP/μg DNA was added in a compatible buffer to the vector preparation and the incubation was continued for 30-60'. If incompatible ends had to be ligated recessive ends were filled in with Klenow fragment as specified by New England Biolabs.

(ii) Purification of vector and insert: vector and insert were separately subjected to electrophoresis (see section 2.5) through a preparative agarose gel containing EtBr. Linearized vector and insert bands were cut out from the gel with sterile scalpels under longwave UV light (365nm) in order to minimize DNA nicking. Then they were transferred to a clean Eppendorf tube. Next, the DNA was isolated from the agarose gel with the QIAquick™ Gel Extraction Kit (Cat# 28704) from Qiagen. The DNA was eluted in 30μl 10mM Tris (pH8.5; provided with the kit).

(iii) Ligation of purified linearized vector and insert: Typically, 50ng purified vector and insert of a 3-fold higher molarity were incorporated into a 10μl ligation reaction (50mM Tris-HCl pH7.5, 10mM $MgCl_2$, 10mM DTT, 1mM ATP, 25μg/ml BSA, 5units T4 DNA ligase). The reaction was incubated at 16°C for 4hrs. A control ligation lacking the insert was routinely carried out to assess the background of recircularized vector molecules. Competent *E.coli* were transformed with 5μl ligation reaction (section 2.2.4).

2.3.4 Small scale preparation of plasmid DNA after ligation

Single bacterial colonies resulting from the transformation were picked and grown in 3ml L-broth supplemented with the appropriate antibiotic for 8-14hrs. For preparation of plasmid DNA from 2ml cultures the alkaline lysis method was employed as described by Sambrook *et al.* (1989). The plasmid was resuspended in 50µl TE. The RNA was digested during the lysis of the bacteria with solution I, and not after resuspension of the DNA. Typically, 4µl (approximately 0.3-0.6µg) were used for restriction analysis and subsequent agarose gel electrophoresis. The restriction enzymes were selected such that the fragment pattern resulting from the recircularized vector and recombinant plasmid were easily distinguishable thus allowing the identification of clones harboring the desired recombinant plasmid ("positive clones"). If the identification of positive clones required the determination of DNA sequence (e.g. after site directed mutagenesis; sections 2.3.6 and 2.3.7), the preparation was extracted with phenol/chloroform (see section 2.3.1) and sent to the sequencing service at the University of Pennsylvania (section 2.3.9). Cultures of positive clones, which were needed for further manipulation or transfection into mammalian cells, were diluted into fresh 100ml L-broth supplemented with the appropriate antibiotic and grown at 37°C overnight for large scale preparation of plasmid DNA.

2.3.5 Large scale preparation of plasmid DNA

DNA was prepared using a modified alkaline lysis method and purified by anion-exchange chromatography (QIAGEN Plasmid Maxi Kit Cat# 12163). The protocol published in the Qiagen manual was followed for large scale preparation of plasmid DNA. After that the DNA pellet was resuspended in 100µl TE and

centrifuged in a microcentrifuge (10'at 13,000rpm). This resulted in the formation of a white precipitate, which is insoluble in TE. The supernatant was transferred to a fresh tube and the centrifugation was repeated. The concentration of the plasmid preparation was determined spectrophotometrically by measuring the absorption of a 200-fold dilution at $\lambda = 260nm$ and applying the formula:

Concentration of double stranded DNA ($\mu g/ml$) = 200 (dilution factor) x 50 x OD_{260nm}.

2.3.6 Transformer™ Site-directed mutagenesis (Clontech)

This is an outline of how the Transformer™ kit works. The exact protocol can be found in the manufacturer's manual.

This site directed mutagenesis is based on the USE (unique site elimination) method. By this method single or multiple specific base mutations as well as deletions or insertions can be introduced. A selection or toggle primer, which mutates a unique restriction site into another anneals to one strand of DNA simultaneously with one or more mutagenic primers. After standard elongation with T4 DNA polymerase followed by ligation with T4 DNA ligase and a primary selection by restriction digest of the mutated unique restriction site the hemi-mutated plasmids are transformed into the BMH71-18 *mutS E coli* strain defective in mismatch repair. Transformants are pooled, and plasmid DNA is prepared from the mixed bacterial population. The isolated DNA is then subjected to a second selective restriction enzyme digestion. Since the mutated DNA lacks the restriction enzyme recognition site, it is resistant to digestion. The parental DNA, however, is sensitive to digestion and will be cut and linearized, making it far less efficient in transformation of bacterial cells. A final transformation with the selectively digested DNA results in highly efficient and

specific recovery of the desired mutated plasmid.

2.3.7 QuikChange® Site-directed mutagenesis (Stratagene)

This is an outline of how the QuikChange® kit works. The exact protocol can be found in the manufacturer's manual. Stratagene's QuikChange® site-directed mutagenesis kit allows site-specific mutation in virtually any double-stranded plasmid. It requires no specialized vectors or unique restriction sites.

This method uses *PfuTurbo*® DNA polymerase and a temperature cycler (PCR machine). The basic procedure utilizes a plasmid with an insert of interest and two complementary oligonucleotide primers containing the desired mutation. The primers are extended during the PCR by *PfuTurbo*® around the whole plasmid. *PfuTurbo*® works with high fidelity and does not displace the mutant oligonucleotide primers. Incorporation of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. After the PCR reaction the product is treated with *Dpn*I. The *Dpn*I endonuclease (target sequence: 5′-Gm6ATC-3′) is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. DNA isolated from almost all *E. coli* strains is dam methylated and therefore susceptible to *Dpn* I digestion. The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue *E. coli* cells. More than 50% of the colonies were positive for the mutation.

The following PCR conditions were used:

94°C	2'10''	
94°C	50''	16x
50°C	2'	
68°C	15'	

94°C	50''	⎫	
65°C	1'	⎬	10x
68°C	15'	⎭	
68°C	15'		

2.3.8 Labeling oligonucleotides

Oligonucleotides were annealed by heating equimolar amounts of compatible oligonucleotides in 100mM NaCl, 1mM $MgCl_2$, 20mM Tris (pH 8) for 5' at 95°C and letting the solution cool down slowly to RT.

The Klenow fragment of *E.coli* DNA polymerase was used to fill-in and label annealed oligonucleotides with 5' overhangs: 5pmol annealed oligonucleotides were filled-in with 2mM dATP, dGTP and dTTP each, plus 2.5µl (α-^{32}P)dCTP (10mCi/ml) by 2.5 units Klenow enzyme in 10mM Tris (pH 7.5), 5mM $MgCl_2$, 7.5mM DTT. After 15' incubation at RT the reaction was filtered through a G25 Sepharose column (Roche, Cat. # 1273949) according to manufacturer's instructions to remove free deoxyribonucleotides.

2.3.9 Sequencing of double-stranded plasmid DNA

Sequencing was performed by the sequencing facility of the School of Veterinary Medicine at the University of Pennsylvania under Dr Jikang Fang. The ABI PRISM® BigDye™ Terminator Cycle Sequencing Kit was used with the ABI PRISM® 377 DNA Sequencer from Applied Biosystems.

2.4 Recombinant DNA Construction

The POU1, POU4, Oct4 and HMG bacterial expression vectors (pET24d(+)) were obtained from Attila Reményi (Reményi *et al.*, 2001; Reményi *et al.*, 2003) and mutated by Transformer™ or by QuikChange® site-directed mutagenesis (sections 2.3.6 and 2.3.7; Chapter 3 and 4). pCGOct1 was obtained from David Denhard and mutated by QuikChange® site-directed mutagenesis for use in transfection assays (Chapter 3). The oligonucleotides used for mutation are listed in section 2.1.4.

The reporter plasmids PORE, $PORE^D$ and $PORE^M$, used in Chapter 3 consist of hexamers of the respective oligonucleotides in front of a thymidine kinase (*tk*) promoter and a *luciferase* gene (*luc*) and are described in Botquin *et al.* (1998) where they are called 6xO, $6xO^{-1}$ and $6xO^{-3}$, respectively.

The following constructs were used in Chapter 5: 6xSDM, 6xsDM reporters (previously named 6xOS and 6xOS-, respectively) and i-opn PCR II were described in Botquin *et al.* (1998). 6xSdM, 6xSDm, 6xsdM, 6xsDm, 6x5SDM, 6x11SDM plasmids were obtained by multimerizing the corresponding oligonucleotides. The 5' overhangs were filled in by Klenow polymerase and fragments of 6 oligonucleotide repeats were precloned into the EcoRV site of pBluescript KS and from there into the HindIII-BamHI sites of the -37*tkluc* plasmid.

2.5 Electrophoresis

2.5.1 Agarose Gel Electrophoresis

Horizontal agarose gel electrophoresis was carried out essentially as described by Sambrook *et al.* (1989). The required weight of agarose was dissolved in 1x TBE by boiling the agarose for several minutes to generate 0.7-2% (w/v) gels. Before pouring the agarose solution into the gel chamber, ethidium bromide was added to a

final concentration of 0.2μg/ml. The gel was left to cool down and solidify at RT and was then submerged in 1xTBE running buffer before removal of the comb. Prior to loading, each sample was mixed with 1/6 volume of a 6x native loading buffer (0.15% bromophenol blue, 0.15% xylene cyanol, 50% glycerol, 60mM EDTA in H_2O) The gel was run at 10-15 V/cm and separated DNA fragments were visualized by UV.

2.5.2 Non-denaturing Polyacrylamide Gel Electrophoresis

Two glass plates, one "notched" plate and one backing plate, were cleaned extensively with ethanol. The size of the plates differed from experiment to experiment. The plates were separated by spacers of 1-1.5mm clamped together and taped. Gels with 5 or 6% acrylamide were routinely used. (150ml of 5% gel: 25ml of 30% acrylamide solution (29:1 acrylamide:bisacrylamide), 30ml 5xTBE, 1.5ml 10% APS, 88.5ml H_2O, 150μl TEMED; the acrylamide and H_2O volumes were adjusted to vary the acrylamide concentration.) The solution was mixed, poured between the glass plates, a comb was inserted and the gel left to polymerase for 20-30'. After polymerization the sealing tape was removed, the plates assembled on the gel apparatus and the wells cleaned with 1xTBE. The gel was exposed to a prerun in 1xTBE at 100V for 30'. Then the samples (5-15μl depending on well size) were loaded and the gel was run at 10V/cm. For some experiments the Novex Mini-Cell system (Cat. # EI0001) from Invitrogen was used with precasted 6% TBE gels (Cat. # EC62652).

2.5.3 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Proteins were separated according to size by the standard Laemmli procedure of one-dimensional gel electrophoresis under denaturing conditions (Laemmli, 1970),

SDS-PAGE. Experiments were carried out using the MGV-202 vertical mini-gel system from C.B.S. Scientific Co., USA.

Two glass plates are assembled using 1mm spacers and the Gel Wrap (used instead of tape). Clamps hold the plates together and hold it upright. To prepare a 10% separating gel 3.325ml acrylamide (29:1 acrylamide:bisacrylamide), 1.25ml 3M Tris (pH8.9), 200µl SDS (10%), 4.875ml H_2O, 100µl TEMED (5%) and 250µl APS (10%) were mixed. For preparation of gels with lower or higher acrylamide concentration the volumes of acrylamide and H_2O were adjusted accordingly. The gel solution is applied to the assembled plates until 1cm beneath the top of the notched plate. 1ml H_2O is added on top of the gel to prevent oxygen from diffusing into the gel and inhibiting polymerization. The gel was left to polymerize for 45'. After polymerization the overlaid H_2O was removed and the stacking gel (825µl acrylamide (29:1 acrylamide:bisacrylamide), 625µl 3M Tris pH 6.7, 100µl SDS (10%), 3.05ml H_2O, 50µl TEMED (5%) and 350µl APS (10%)) was added on top of the running gel. A 1mm comb was inserted into the stacking gel and it was left to polymerize for 15'. Samples were loaded after the gel was mounted on the gel chamber and after running buffer (23mM Tris pH 8.4, 190mM glycine, 0.2% SDS) was added. The proteins were run at 100V through the stacking gel and at 180V through the separating gel.

2.6 Handling of Proteins

2.6.1 Detemination of protein concentration

The Bio-Rad protein assay (Cat# 500-0006), based on the Bradford dye-binding procedure, was used to determine protein concentration. Protein standards ranging from 1-20µg were prepared with BSA. Standards and proteins were brought to 20µl. To these 800µl H_2O and 200µl Bio-Rad protein assay solution were added.

After 15' incubation at RT the absorption was measured with a spectrophotometer at 595nm in plastic cuvettes. Absorption values of the BSA standards were used to calculate the protein concentration of the samples. In addition, protein concentrations were estimated by running SDS-PAGE gels (section 2.5.3) with known amounts of protein along with newly purified protein and concentrations were estimated by comparison using Coomasie Blue staining (section 2.6.2).

2.6.2 Coomassie staining of protein gels

The SDS gel was submerged in Coomassie Blue staining solution (50% methanol, 12% acetic acid, 0.05% Coomassie Brilliant Blue R250 (w/v)) so that it was covered completely and incubated o/n at RT shaking at 65rpm. The staining solution was then replaced by destaining buffer (50% methanol, 12% acetic acid). The gel was washed 3-4 times for 20-40' with the destaining buffer until blue colored protein bands became visible against a clear background within the gel. The gel was then dried on Whatman paper (grade 3) in the Slab Gel Dryer from Savant (SGD5040) coupled to the Universal Vacuum System (Savant, UVS400) for 1hr at 80°C for long term storage.

2.6.3 Western blotting

The following protocol was used to transfer proteins from an SDS-PAGE gel (section 2.5.3) to BA23 nitrocellulose membrane (Schleicher Schuell; pore size 0.2μm) using the EBU-202 transfer chamber from C.B.S. Scientific Co., USA. All required components were soaked in Western transfer buffer (25mM Tris (pH 8.3), 192mM glycine, 20% (v/v) methanol) before use and assembled in the following order: a plastic sponge pad, 3 sheets of Whatman paper (grade 3), the polyacrylamide gel, a piece of nitrocellulose membrane, a second group of 3 sheets of Whatman

paper, and then another plastic sponge. Before assembly air bubbles that were trapped between the gel and the nitrocellulose were removed by rolling a serological pipette over the surface of the nitrocellulose and following Whatman paper. The entire assembly was placed into the plastic holder and slid into the transfer chamber filled with Western transfer buffer so that the nitrocellulose was facing the anode and the gel was facing the cathode. Proteins were transferred to the nitrocellulose by applying a constant voltage 130mA o/n at 4°C. During this time the transfer buffer was stirred continuously with a magnetic stirrer bar.

The transfer efficiency was analyzed by incubating the nitrocellulose membrane in Ponceau-S solution (composition: 2% Ponceau-S (w/v) in 30% trichloroacetic acid, 30% sulfosalicylic acid) for a couple of minutes and rinsing it with ultrapure H_2O until the protein bands were clear and there was minimal background staining of the membrane.

After all Ponceau-S staining was removed by washing the nitrocellulose membrane with ultrapure H_2O, the membrane was transferred into blocking solution (PBS, 5% fat free milk powder, 0.05% Tween-20) either for 2hrs at RT or o/n at 4°C. The primary antibody was appropriately diluted (about 1:1000) in fresh blocking buffer and incubation was continued with gentle agitation for 1hr at RT. The membrane was washed 4 times in PBS + 0.05% Tween-20 (5' each) to remove any unbound antibody. Then it was incubated with a secondary horseradish peroxidase coupled anti-rabbit antibody (1:1500) in fresh blocking solution for 1hr and washed as before. Next, the membrane was submerged in ECL™ Western blotting detection solution (reagents 1 and 2 freshly mixed 1:1; Amersham Pharmacia) for 1'. ECL™ Western blotting is a light emitting non-radioactive method for detection of immobilized specific antigens, conjugated with horseradish peroxidase-labeled

antibodies. The membrane was exposed to Kodak ML Biomax™ film for 30 seconds and developed for visualization of reactive protein bands. The developed film determined how strong the ECL reaction was and the length of exposure (30''-30') for the final image.

2.6.4 Purification of proteins

POU1, HMG and their derivatives were isolated in the following way. All other proteins used were isolated similarly by collaborators (Reményi *et al.*, 2001; Reményi *et al.*, 2003; Botquin *et al.*, submitted).

E. coli of the BL21-CodonPlus(DE3)-RIL strain were transformed with the pET24d(+) vector containing the His-tagged POU domain of Oct1, the HMG domain of Sox2 or mutants thereof. After growth on an agar plate supplemented with Kanamycin o/n, one colony was used to inoculate 6ml LB medium + Kanamycin and incubated at 37°C o/n (preculture). The next morning 200ml LB + Kanamycin were inoculated with 4ml of the preculture (1:50 dilution). At an OD_{600} of 0.4-0.6 protein expression was induced by adding 1mM IPTG. After 2-4hrs of protein expression (shaking at 30°C) the cells were pelleted (15' at 3600rpm = 2550 rcf in Thermo Forma Multi RF Centrifuge with the swinging bucket rotor #5808947). The pellet was resuspended in ice cold 50ml PBS to wash the bacteria from the media, and pelleted again (15' at 3600rpm = 2550 rcf). This pellet was taken up in 10ml lysis buffer (500mM NaCl, 20mM Hepes pH7.6, 10mM imidazole, 5mM β-ME, 0.1% IGEPAL-CA-630) and transferred to a Falcon tube. The sample was then sonicated for 5' on ice (Ultrasonic processor CP130 (Cole Parmer Cat. # EW-04714-20); amplitude: 70%, pulser: 10", using a 6mm titanium probe and tip (Cat. # EW-04712-14)). After the cells were completely lysed the solution was centrifuged at 10000rcf for 30'

(SLC-250T rotor in Sorvall Super 21T centrifuge). Batch purification was performed using Ni-NTA agarose resin (Qiagen #1000 630). 800µl of the resin were added (previously equilibrated with lysis buffer) to the supernatant and incubated at 4°C rotating at 50rpm for 1-12hrs. Next the bacteria-resin solution was transferred to Polyprep chromatography columns (BioRad #731-1550). The resin was washed with 10ml lysis buffer, then 5ml NaCl solution (1M NaCl, 20mM Hepes pH7.6, 10mM imidazole, 5mM β-ME, 0.1% IGEPAL-CA-630) followed by 5ml 35mM imidazole solution (150mM NaCl, 20mM Hepes pH7.6, 35mM imidazole, 5mM β-ME, 0.1% IGEPAL-CA-630). The protein was eluted in 800µl elution buffer (150mM NaCl, 20mM Hepes pH7.6, 300mM imidazole, 5mM β-ME, 0.1% IGEPAL-CA-630 and dialysed against dilution buffer (150mM NaCl, 20mM Hepes pH7.6, 1mM EDTA, 5mM β-ME, 0.1% IGEPAL-CA -630) o/n at 4°C.

2.7 Cell Lines and Cell Culture

2.7.1 Cell culture media, reagents and cell lines

Cell culture reagents were prepared and all tissue culture manipulations were carried out in sterile conditions in a standard laminar flow hood. Dulbecco's Modified Eagle's Medium (DMEM), supplements (L-Glutamine, penicillin/streptomycin) and reagents (Trypsin-EDTA, PBS) were purchased from Invitrogen, USA. Defined Fetal Calf Serum (FCS) was obtained from HyClone, USA. The 293, F9 and R1 cells were propagated in an incubator maintaining a constant temperature of 37°C and a humidified, 5% CO_2 atmosphere. Cells were grown in cell culture dishes (15cm diameter, 24-well dishes). The cell lines utilized are indicated below:

Cell line:	Cell type:	Obtained from:
293	Human, adenovirus-transformed line derived from embryo kidney	P. Gruss, MPI, Göttingen, Germany
BJA-B	Human, B-cell lymphoma derived	P. Gruss, MPI, Göttingen, Germany
F9	Mouse, embryonal carcinoma (EC)	ATCC, USA
R1	Mouse, embryonal stem cells (ES)	Nagy, Mount Sinai Hospital, Toronto, Canada

The 293 and F9 EC cells were maintained in DMEM supplemented with 10% FCS, penicillin/streptomycin (P/S) and L-Glutamine (L-Glu) (per liter: 880ml DMEM, 100ml FCS, 10mls 100x P/S, 10mls 100x L-Glu). R1 ES cells were cultured in DMEM (0.45% glucose (wt/vol)), 15% FCS, 100 μM β-ME, penicillin/streptomycin (P/S), L-Glutamine (L-Glu) and in the presence of 1000 U/ml leukemia inhibitory factor (LIF, GIBCO-BRL). Murine EC and ES cells were grown on 0.1% gelatin-coated tissue culture plates. For methods concerning BJA-B-cells see Lins *et al.* (2003).

2.7.2 Passage of cells

Cells were passaged using conventional cell culture techniques. Briefly, when reaching 80-90% confluency, cells were washed with sterile PBS and incubated in 3ml trypsin-EDTA (for 15cm plate) for 1'. Trypsinization was stopped by the addition of medium and a single cell suspension was created by pipetting the suspension up and down. An appropriate volume was then transferred to dishes with fresh medium. For long term storage cells were frozen in 40% DMEM, 50% FCS, 10% DMSO. To bring cells into culture from a frozen stock the suspension was thawed quickly and

directly added to the appropriate amount of pre-warmed supplemented medium in a culture dish.

2.7.3 Transient transfection of 293, F9 and R1 cells

Cells were transfected with the non-liposomal FuGENE 6 Transfection Reagent from Roche (Cat. # 1 814 443). FuGENE 6 Transfection Reagent is a multi-component lipid-based transfection reagent that complexes with and transports DNA into the cell during transfection.

Cells were fed with supplemented medium without P/S, since this may adversely affect transfection efficiency, 2hrs before the transfection. Per well of a 24-well dish, 1μg DNA was transfected. For this, 2μl FuGENE 6 were added to 100μl DMEM followed by addition of 1μg DNA, always consisting of 100ng pcmvβgal to measure galactosidase activity as an internal standard and 200ng reporter plasmid. When appropriate different amounts of expression vectors were added. The suspension was incubated at RT for 15' and then added to the cells dropwise. After 48hrs cells were harvested and whole cell extracts were prepared (section 2.7.5).

2.7.5 Preparation of whole cell extracts from transfected cells

Cells in 24-well dishes were lysed in the wells. The cells were washed with 500μl ice-cold PBS twice. Then 150μl lysis buffer (250mM Tris pH7.8, 1mM DTT protease inhibitors Aprotinin (2μg/ml) and Leupeptin (5μg/ml) (Roche) were added to each well. The cells were freeze-thawed three times, placing the plates alternately in liquid N_2 and a 37°C water bath. The lysates could then be stored at -80°C or used directly. Whole cell extracts prepared in this manner were used for luciferase, β-galactosidase and electrophoretic mobility shift assays.

2.7.6 Enzymatic assays of extracts from transfected cells

Luciferase Assay

The luciferase assay measures the activity of the luciferase reporter constructs in the transfected cells and was performed in the Mediators PhL luminometer (Mediators Diagnostika, Austria; Cat. # G010001.01)

150µl of freshly prepared Mix A (for 30samples: 4.9ml Glycylglycine pH7.8, 50µl ATP pH7.5 (0.2M), 50µl $MgSO_4$ (1M)) were added in each well of a FluoroNunc 96-well Module (MaxiSorp Surface, Nunc #437591). 20µl whole cell extract were added to the wells. Freshly prepared Mix B (for 30 samples: 4ml Glycylglycine pH7.8, 1ml D-luciferin solution (5mg/16.5ml of the above Glycylglycine buffer) was put in a light protected tube. After the machine was primed (see user manual) the Mix B tube was hooked up to the Mediators PhL luminometer and the FluoroNunc plate was placed in the machine. Luciferase activities were measured with the following settings: 30"integration time, 2" lag time, 50µl injection volume for Mix B. The luciferase measurements were normalized for (divided by) the corresponding β-galactosidase value.

β-galactosidase assay

To determine the relative amount of β-galactosidase protein in the whole cell extract of transfected cells a colorimetric enzyme assay was employed. 10-20µl extract were added to 60µl solution I (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgCl_2$, 50mM β-ME) in a transparent 96-well plate. 20µl solution II (2mg/ml ONPG (β-galactosidase enzyme substrate) in Solution I) were added and samples were incubated at RT. Conversion of ONPG by β-galactosidase results in a color change of the solution from transparent to yellow and thus allows

spectrophotometric quantification of β-galactosidase present in the sample. When the color in the wells turned obviously yellow, 50µl solution III (1M Na_2CO_3) were added to terminate the reaction. The optical densities at 420nm were read in the Mediators PhL luminometer. As β-galactosidase expression was used as an internal standard in the transfection experiments, i.e. each well in a transfection series was transfected with the same amount of β-galactosidase expression vector, differences in β-galactosidase activity were assumed to reflect differences in transfection efficiency and the reporter (luciferase) values were corrected accordingly.

2.8 DNA/protein binding assays

2.8.1 Electrophoretic Mobility Shift Assays (EMSAs)

For experiments in Chapter 3:

Approximately 30ng POU1 (wild-type or mutant) protein was incubated with radiolabeled oligonucleotides (see section 2.3.8) in the presence or absence of 10ng chemically synthesized OBF1 polypeptide (see section 2.1.5). The following protein-DNA binding buffer was used for the reaction: 25 mM Tris pH 8.0, 100 mM NaCl, 0.5 mM EDTA, 0.04% Triton-X, 10% glycerol, 10 mM DTT, 10ng/µl poly[d(I-C)].

For experiments in Chapters 4 and 5:

Combinations of approximately 25ng POU1, 100ng Oct4, 25ng POU4, 100ng Pax6 and 50ng HMG or indicated amounts of F9 whole cell extract (see section 2.7) were incubated with the appropriate labeled oligonucleotides in the protein-DNA binding buffer (10mM Tris, pH 8.0, 150mM NaCl, 0.05% Triton-X, 5 µg/µl BSA, 50ng/µl salmon testes DNA, 10% glycerol, 10mM DTT, 10ng/µl poly[d(G-C)]) and complex formation was analyzed by EMSA. Titration experiments were carried out

under similar conditions using the same amount of Oct4 (100ng) with increasing amounts of HMG domain of Sox2 (0, 6.25, 12.5, 25, 50, and 100ng) per binding reaction.

The EMSA gels were dried on Whatman paper (grade 3) in the Slab Gel Dryer from Savant (SGD5040) coupled to the Universal Vacuum System (Savant, UVS400) at 80°C for 45'. The dried gel was then exposed to Kodak X-OMAT™ Blue XB-1 film for 15-90'. The radioactive signal was enhanced by an NEN intensifying screen at RT.

2.8.2 Off-rate EMSA and dissociation rate constants

For the off-rate experiment (figure 3.7), a 500-fold excess of unlabeled oligonucleotide was added to the reaction. Aliquots of the reaction mixture were loaded onto the gel at indicated time intervals.

The dissociation rates of Oct4 and Sox2 containing complexes were determined in Chapter 5. The retarded bands were quantified using the Fujifilm-2000 phosphorimager, Basread and Aida version 2.0 programs. Percentage of bound complex at a specific time (bound protein-DNA complex at time-x)/($\sum$ of total bound protein-DNA complex at time-x) was then plotted as a function of time as a semi-logarithm graph. The half-life was derived from the dissociation curve as the time at which 50% of the DNA-protein complex was still intact and had not dissociated yet. From the slopes, the dissociation rate constants (k_d) of the dimension 1/sec were calculated with the equation $k_d = \ln 2 / T_{50\% \text{ complex bound}}$.

Chapter 3

OBF1 enhances transcriptional potential of Oct1

3.1 Abstract

The POU transcription factors Oct1 and Oct2 bind to DNA in various monomer and dimer configurations. Depending on the DNA sequence they bind to, the dimers are arranged in configurations that are either accessible (PORE sequence) or inaccessible (MORE sequence) to the B-cell specific cofactor OBF1 (OcaB, Bob1). As shown previously, the MORE and related sequences (such as the Heptamer/Octamer motif) are found in immunoglobulin heavy chain promoters. In this chapter I show that the expression of *Osteopontin*, which contains a PORE sequence in its enhancer region, depends on the presence of OBF1 in B-cells. OBF1 alleviates DNA sequence requirements of the Oct1 dimer on PORE-related sequences *in vitro*. Furthermore, OBF1 stabilizes POU dimer-DNA interactions and overrides Oct1 interface mutations, which abolish PORE-mediated dimerization without OBF1. The data presented here indicate that the PORE-type Oct1 or Oct2 dimer, rather than the monomer, is the primary target of the cofactor OBF1. Based on this biochemical data, I propose a mode of OBF1-Oct1 dimer interaction, indicating a novel arrangement of the subdomain connectivities.

3.2 Introduction

Specificity in the transcriptional regulation of gene expression is necessary to enable the correct temporo-spatial expression pattern during development. The combination of multiple factors represents an efficient adaptation to integrate different signal pathways and to coordinate cell type and cell cycle specificity. To this end, transcription factors bind to DNA directly and assemble with each other and with coactivators and/or corepressors. This leads to the formation of specific

transcriptional complexes with distinct characteristics based on the DNA-binding sequence of the regulatory regions and the particular factors involved.

The POU family of transcription factors is involved in the transcriptional regulation of a wide array of ubiquitous and tissue-specific genes. Oct1, often regarded as the prototype POU factor, is broadly expressed. Oct2 and Oct4 are exemplary POU genes with a narrow expression profile. Oct2 is mainly found in cells of the lymphoid system and Oct4 is limited to the mammalian germline, including stem cells of the early embryo and germ cells (reviewed in Ryan and Rosenfeld, 1997).

Members of the POU transcription factor family share a conserved bipartite DNA-binding domain called POU domain, containing the modular POU-specific domain (POU_S) and POU-homeodomain (POU_H). The subdomains are connected by a flexible linker, variable in sequence and length (15-56 amino acid residues). POU factors bind to DNA elements, such as the octamer motif (ATGCAAAT), as monomers (Staudt *et al.*, 1986; Schöler *et al.*, 1989b). More recently, POU factors have been shown to homo- and heterodimerize on specific DNA motifs (Jacobson *et al.*, 1997; Botquin *et al.*, 1998; Rhee *et al.*, 1998; Scully *et al.*, 2000; Tomilin *et al.*, 2000; Reményi *et al.*, 2001). The Oct-factor subgroup binds two of these octamer-related sequences – the PORE (Palindromic Oct factor Recognition Element: ATTTGAAATGCAAAT) and the MORE (More palindromic Oct factor Recognition Elements: ATGCATATGCAT). The PORE was first identified as an Oct4 binding sequence in the first intron of the *osteopontin* (*OPN*) gene in embryonic carcinoma (EC) cells (Botquin *et al.*, 1998). Homo- and heterodimers of various POU factors, including Oct1, Oct2, Oct4 and Oct6, can assemble on the PORE *in vitro*. Dimerization on the PORE mediates strong transcriptional activation supporting the

notion of dimer formation *in vivo* (Botquin *et al.*, 1998). MOREs and related sequences (e.g., the Heptamer/Octamer motif) are found in immunoglobulin heavy chain promoters (V_H). Oct family members can also bind cooperatively as homo- and heterodimers on the MORE *in vitro* (Tomilin *et al.*, 2000).

Furthermore, POU proteins form heterodimers with transcription factors of other families. Several POU factors have been shown to interact with members of the HMG family, such as Sox2, *in vitro* (Yuan *et al.*, 1995; Botquin *et al.*, 1998; Nishimoto *et al.*, 1999; Ambrosetti *et al.*, 2000; Chapter 4, Chapter 5) or to assemble with cofactors (Sauter and Matthias, 1998; Scully *et al.*, 2000). The most extensively studied cofactor is OBF1, which is expressed in lymphoid cells and interacts specifically with Oct1 and Oct2 monomers and dimers (Luo *et al.*, 1992; Gstaiger *et al.*, 1995; Luo and Roeder, 1995; Strubin *et al.*, 1995; Tomilin *et al.*, 2000). The interaction with the dimer is dependant on the conformation of the POU/DNA complex, as OBF1 can only bind to PORE-mediated Oct1 dimers but not to those on MOREs. The recently solved crystal structures of the POU domains of the Oct1 dimer bound to the MORE and PORE revealed the structural basis for this selectivity of OBF1. The two POU dimers adopt two different configurations, using different surface patches to dimerize on each element (Tomilin *et al.*, 2000, Reményi *et al.*, 2001). As a result, the same amino acids that are available to interact with OBF1 in the PORE configuration (Chasman *et al.*, 1999) form part of the POU_S-POU_H dimer interface on the MORE, precluding an interaction with OBF1.

Osteopontin (OPN) is expressed in various tissues (Denhardt *et al.*, 1995) and the immune system (Weber and Cantor, 1996 and O'Regan and Berman, 2000), e.g., in T-lymphocytes and activated pro-B-cells (Lin *et al.*, 2000). Secreted OPN stimulates B-cells to produce immunoglobulins and, in conjunction with an

unidentified 14kD peptide, to proliferate (reviewed in Weber and Cantor, 1996). Moreover, B-lymphocytes have been shown to play a role in new bone formation in which OPN is also involved (Marusic *et al.*, 2000). One common theme that has emerged from several of these studies is that OPN is involved in cell migration and adhesion.

In this chapter, I first demonstrate a regulatory link between POU proteins, OBF1 and OPN expression in B-cells. I then describe the effect of an N-terminal OBF1 peptide (sufficient for protein and DNA binding) on the PORE-mediated dimerization of the POU domain of Oct1 (POU1). The transfection experiment with BJA-B-cells reveals that OPN expression in lymphoid cells depends on OBF1 and that the PORE element is active in B-cells. Consequently, lymphoid cells provide (i) an environment conducive to the differential expression of genes via different conformations of the same transcription factor and (ii) cofactor recruitment, which are both dictated by the DNA sequence. Using Oct1 interface mutants that specifically inhibit dimerization on the PORE, I show that OBF1 can compensate for the loss of the PORE-specific dimer interface. Besides overriding the effect of the mutations, OBF1 alleviates DNA sequence requirements by clamping the dimer to the DNA, significantly stabilizing the protein-DNA complex. My results suggest that an arrangement of the POU subdomains is adopted by the Oct1-OBF1 complex on the PORE that is different from that proposed for Oct1 alone.

3.3 Results

3.3.1Osteopontin is a target gene of OBF1

So far, MOREs have been shown to mediate transcriptional activation in B-cells while POREs are active in pluripotent embryonal cells (Botquin *et al.*, 1998; Tomilin *et al.*, 2000). To determine whether both regulatory elements can be functional in the same cell type *in vivo*, my collaborators and I examined if POREs are active in B-cells. To this end, I studied the PORE located in the first intron of *OPN*, which is known to regulate *OPN* expression in ES and EC cells (Botquin *et al.*, 1998). First, the levels of *OPN* transcripts were analyzed by Northern blot analysis (performed by Steffen Massa, FMI, Basel, Switzerland; for method see Lins *et al.*, 2003) to compare the mRNA of normal and OBF1-deficient splenocytes, stimulated *in vitro* (figure 3.1A). *OPN* transcription was far stronger in wild-type splenocytes than in OBF1-deficient cells. We concluded that *OPN* is a target gene of this transcriptional coactivator *in vivo*. However, indirect stimulation of *OPN* via activation of another gene cannot be excluded.

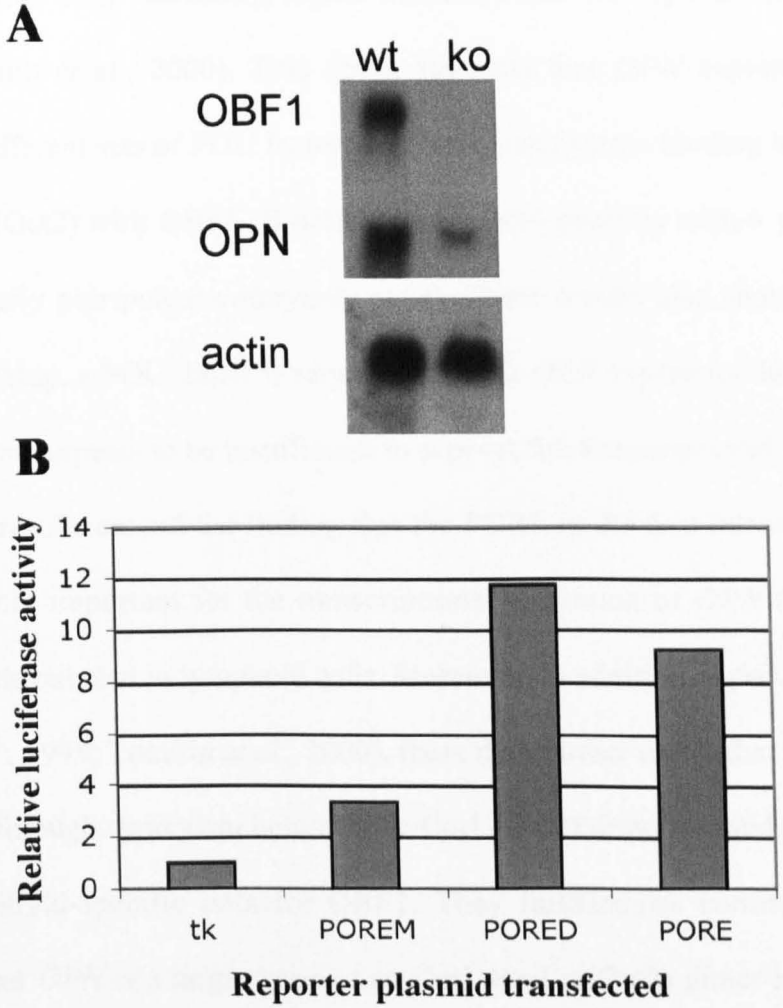
To determine whether OBF1 has the potential to stimulate the *OPN* gene directly, we analyzed several PORE variants in B-cells. For this purpose, three different PORE-variant reporter plasmids were transfected into BJA-B-cells known to express high levels of OBF1 protein (performed by Alexey Tomilin, UPENN, USA; for method see Lins *et al.*, 2003). Hexamers of the PORE sequence were cloned in front of the thymidine kinase minimal promoter (*tk*) driving the *luciferase* gene. The mutation in the $PORE^D$ restricted binding to the POU dimers *in vitro* while that in the $PORE^M$ restricted binding to POU monomers (Botquin *et al.*, 1998). In F9 EC cells, both the PORE and $PORE^D$ were more active than $PORE^M$, which was only slightly more active than the *tk*-promoter alone (Botquin *et al.*, 1998). In B-cells, reporter

Figure 3.1

Osteopontin is regulated by OBF1 in lymphoid cells

(**A**) Analysis of *OPN* expression in wild-type (wt) and OBF1 -/- (ko) splenocytes. Cells were stimulated *in vitro* and total RNA was analyzed by Northern blotting. Osteopontin (OPN) transcripts were found at high levels in wild-type splenocytes. OBF1 -/- splenocytes show reduced OPN transcript levels. Actin: control for comparable mRNA levels in both samples; OBF1: control to show presence of transcripts in wild-type cells and absence in knock-out cells.

(**B**) Comparison of enhancer activities of PORE-derived elements in transient transfection experiments. BJA-B-cells were transfected with different luciferase reporter plasmids (X axis). Y axis: activation of transcription, expressed as relative luciferase activities. The *tk* minimal promoter served as a control. BJA-B-cells naturally express OBF1. $PORE^D$ and $PORE^M$ are derivatives of the PORE, to which only the POU dimer and monomer can bind, respectively.


A
wt
ko
OBF1
OPN
actin
B
Relative luciferase activity
14
12
10
8
6
4
2
0
tk
POREM
PORED
PORE
Reporter plasmid transfected

activity compared to that of the *tk* reporter alone, was about 9-fold higher for the PORE, 12-fold higher for the $PORE^D$, and 3-fold higher for the $PORE^M$ (figure 3.1B). These levels of increased reporter activity are comparable to those obtained in cotransfection with OBF1 experiments using 293 cells. OBF1 stimulates PORE activity, with the $PORE^D$ mediating higher transcriptional activity than the PORE and $PORE^M$ (Tomilin *et al.*, 2000). This result suggests that *OPN* expression can be activated by different sets of POU factors and their coactivators binding to the PORE, e.g., by Oct1 (Oct2) with OBF1 (B-cells) and by Oct4 possibly with a yet unknown coactivator (early pluripotent embryonic cells). These results also show that in this experimental setup, a POU dimer is required for high *OPN* expression levels whereas a POU monomer appears to be insufficient to support full transcriptional activation.

These results extend the finding that the PORE in the first intron of the *OPN* gene is not only important for the transcriptional activation of *OPN* in pluripotent embryonal cells but also in lymphoid cells. Moreover, in addition to previous findings (Botquin *et al.*, 1998; Tomilin *et al.*, 2000), these data further imply that the activation is modulated through synergism between the Oct1 (Oct2) dimer formed on this PORE and the lymphoid-specific cofactor OBF1. They furthermore confirm my earlier hypothesis that *OPN* is a target gene of an Oct1 (and/or Oct2) dimer in conjunction with OBF1 in lymphoid cells.

3.3.2 Identification of potential PORE-interface mutants

The above-mentioned experiments suggest that an Oct dimer together with OBF1 plays a critical role in regulating *OPN* expression in lymphoid cells. To analyze the molecular interaction between these proteins, a set of specific mutations were introduced into the POU domain of Oct1 (POU1), based on the information provided

by the crystal structure of the PORE-mediated POU1 dimer (Reményi *et al.*, 2001). The non-overlapping nature of the MORE and the PORE dimerization interfaces allowed us to design mutants that were able to selectively affect one type of POU dimer formation while leaving the other intact (figure 3.2A and published in Reményi *et al.*, 2001).

In the crystal structure of the POU-PORE complex, the POU_S-POU_H dimer interface is formed by three types of interactions: (i) POU_S-POU_H salt bridges (D29-K104 and K22-E109); (ii) specific hydrogen bonds of R20 (POU_S) and S107 (POU_H) to a common phosphate group in the minor groove of the DNA; and (iii) van der Waals interactions within the POU_S-POU_H interface (by Q18, I21, K22, K104, S107 and E109) (figure 3.2B, C). In order to affect dimer formation on the PORE element, amino acid residues that play a prominent role in the PORE-type interface (POU_S: Q18, I21 and K22; POU_H: K104, S107 and E109) were mutated either into small amino acids (A, G or S) to remove side-chain specific interactions, or into amino acids containing bulky or charged side-chains (Y, D or E) to cause steric clashes or electrostatic repulsion within the interface. Two of these mutations are particularly interesting since they imitate phosphoserines (m12, S107D; m13, S107E; Maciejewski *et al.*, 1995). The list of the Oct1 POU domain mutants is shown in figure 3.2D.

The POU1 dimer-interface mutants were expressed in *E. coli* and their protein-DNA and protein-protein interactions were analyzed by electrophoretic mobility shift assays (EMSAs). First, monomer formation on the canonical octamer motif was tested using an oligonucleotide derived from the Igκ promoter (Bergman *et al.*, 1984; figure 3.3 odd lanes). Two mutants – m3 (I21Y) and m11 (S107A) – had only little effect on DNA binding. For m3 monomer binding was only affected

Figure 3.2

POU_S-POU_H interface in the PORE-type dimer

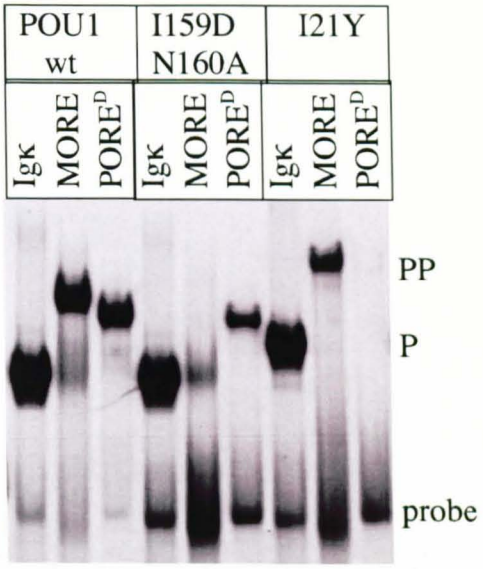
(A) Interface mutants selectively disrupt the MORE- and PORE-type Oct1 dimer. Amino acid residues from the MORE- or PORE-type POU_S-POU_H interfaces were mutated to disrupt the respective dimer (MORE-interface mutant: I159D, N160A; PORE-interface mutant: I21Y). According to the EMSA, the interface-specific mutations do not interfere with monomer binding to the Igκ octamer site but selectively disrupt the MORE- or PORE-type complex formation. WT, wild-type POU domain of Oct1; Igκ, oligonucleotide containing the octamer motif from the immunoglobulin kappa chain promoter; P, POU monomer/DNA complex; and P/P, POU dimer/DNA complex. The unmutated POU1 and I159D,N160A mutant proteins exhibit a slightly higher mobility than the I21Y mutant protein, which has an additional 3kD histidine tag. The tag slows down migration in the gel, but does not affect interaction with the DNA.

(B) Representation of the POU1 dimer on the PORE. The coordinates were taken from the crystal structure published by Reményi *et al.* (2001). One POU1 molecule is colored in purple; the other in green.

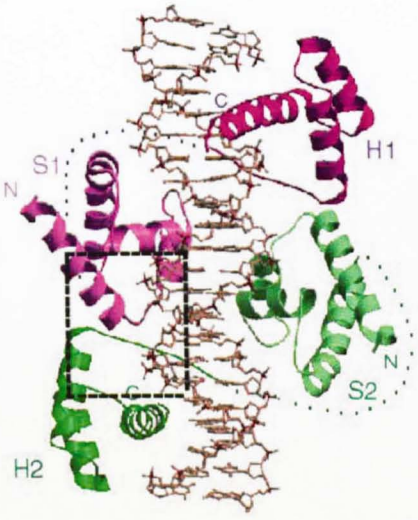
(C) Close-up of the POU_S-POU_H interface focusing on the intermolecular protein-protein interactions. All amino acid residues mentioned are predicted to play a role in forming the dimer interface. POU_S is colored in purple, POU_H in green. The critical residues are indicated in yellow and the DNA is colored grey. Two-digit amino acids are part of POU_S; three-digit amino acids belong to POU_H.

(D) List of all mutants used in this study.

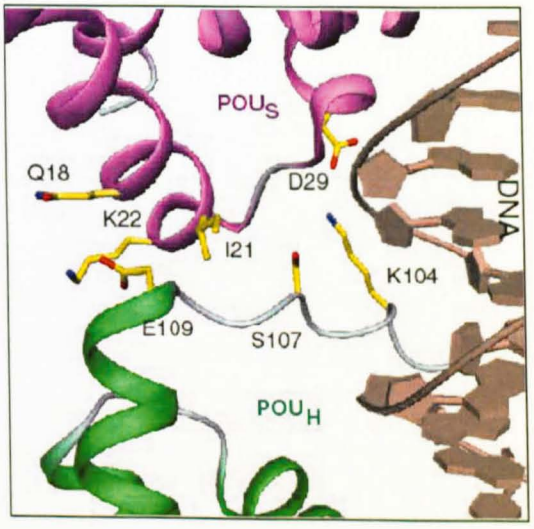
A



B



C

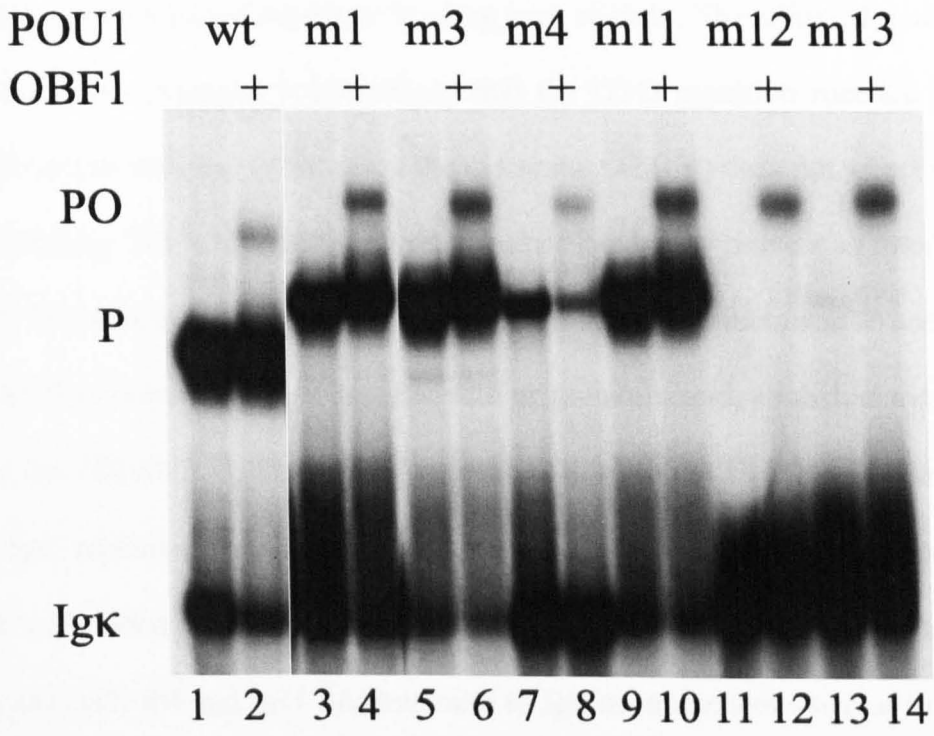


D **POU1 Mutants**

m1: Q18A, I21A, K104S, E109A
m3: I21Y
m4: I21G, K22A
m11: S107A
m12: S107D
m13: S107E

Figure 3.3

EMSA showing different effect of POU1 mutants on monomer binding to Igκ, an octamer-containing promoter. Odd lanes: Binding pattern of POU1 and its mutants, respectively, on the octamer site. Even lanes: Binding pattern of OBF1 plus POU1 or its mutants on the octamer site. P: POU1/DNA complex; PO: POU1/OBF1/DNA complex.


POU1 wt m1 m3 m4 m11 m12 m13
OBF1 + + + + + + +
PO
P
Igκ
1 2 3 4 5 6 7 8 9 10 11 12 13 14

slightly, as I21 is not involved in DNA binding and this residue is far away from the double helix (figure 3.2B+C; Klemm *et al.*, 1994). In m11, the serine at position 107 (S107) was changed to alanine – a mutation that has no effect on DNA binding, as this residue is not involved in POU monomer contacts to Igκ (Klemm *et al.*, 1994). In m1, which was generated to preclude dimerization between POU_S and POU_H, residues Q18, I21, K104 and E109 were mutated into alanines and one serine (see figure 3.2D), which affected monomer binding very slightly. The effect of mutations in m4 was slightly stronger. It is unlikely that the I21G mutation resulted in the weaker interaction with the DNA, since the m3 mutant (I21Y) does not affect DNA-monomer binding. The K22A mutation, on the other hand, may prevent an interaction with E109. In marked contrast, the phosphoserine mimicking mutants m12 and m13, in which S107 is mutated into an aspartic acid or glutamic acid, abolished monomer binding to Igκ (figure 3.3). Since S107 is located close to the DNA helix when it is bound to Igκ, replacing it with the bulky and negatively-charged residues probably interferes with DNA binding due to electrostatic repulsion and steric clash. In summary, m1, m3, m4 and m11 mutants bind to Igκ as monomers whereas m12 and m13 do not.

Several studies had previously described a clamping effect of OBF1 on the POU1 monomer to DNA (Babb *et al.*, 1997; Sauter and Matthias, 1998). To assess whether OBF1 can overcome the compromised DNA binding activities of some of these POU1 mutant proteins, I compared their DNA binding in the presence and absence of the OBF1 peptide used for solving the Oct1/OBF/DNA crystal structure (figure 3.3; Chasman *et al.*, 1999). None of the mutations had a negative effect on the POU1-OBF1 interaction. OBF1-assisted binding was equally weak for all POU1 proteins that showed normal DNA binding and did not correlate with their intensity of

monomer binding to Igκ DNA (compare even and odd lanes). Furthermore, m12 and m13, which cannot bind to Igκ alone, also formed an OBF1-containing complex. These latter results indicate that OBF1 enables Oct1 binding to Igκ DNA when monomer binding is severely compromised.

3.3.3 Structural Basis of OBF1 Interaction with the Oct1/PORE Complex

The crystal structure of the POU1/PORE dimer complex (Reményi *et al.*, 2001) suggests the way OBF1 could bind to the PORE mediated Oct1 dimer. This structure shows how the deviations of the PORE sequence from palindromic symmetry lead to differences in the geometric parameters of the two DNA half-sites which in turn result in an asymmetric dimer arrangement. Superposition of the two half-site POU_S-POU_H arrangements of the PORE dimer onto the Oct1/octamer/OBF1 complex (Chasman *et al.*, 1999) shows that only one of them, covering the octamer-like half-site of the PORE (ATTTGAAATGCAAAT), matches the OBF1 binding site geometry of the Oct1/octamer/OBF1 complex (figure 3.4A). Within the other POU_S/POU_H/DNA associate from the nonoctamer half-site, the POU_H domain is displaced with respect to the arrangement of the Oct1/octamer motif complex as a reference. In this conformation, the C-terminus of the third POU_H helix, which is critical for POU/OBF1/DNA complex formation (Chasman *et al.*, 1999), would be too far away to be involved in any specific interactions with the coactivator. These structural data are validated by a biochemical experiment showing that only one OBF1 coactivator molecule binds to the Oct1 PORE dimer (figure 3.4B). In essence, these data demonstrate that regulation of Oct1 by binding of the coactivator OBF1 not only depends on its overall dimer arrangement but also on the specific geometry of the DNA element (figure 3.4C).

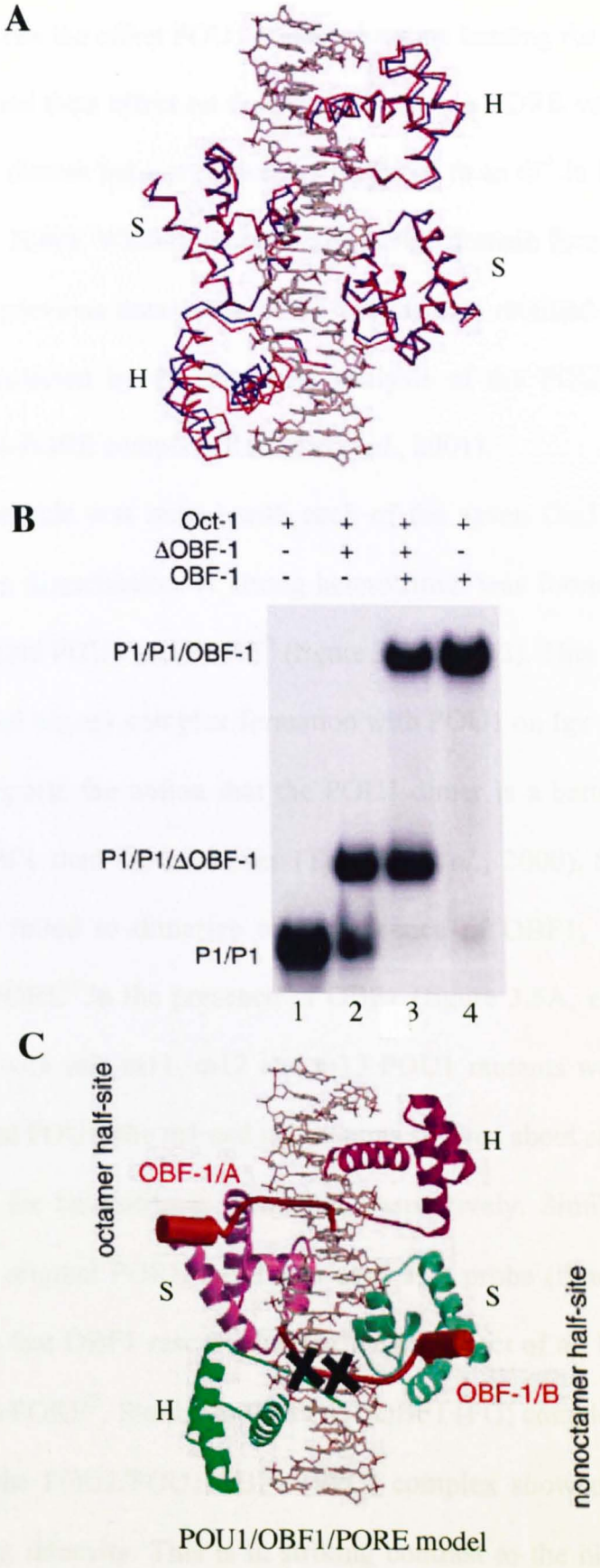
Figure 3.4

Model of Oct1/PORE/OBF1 ternary complex

(**A**) Superposition of the Oct1 POU domain from the octamer/Oct1/OBF1 peptide ternary complex (Chasman *et al.*, 1999) with the two POU domain monomers from the Oct1/PORE crystal structure (see text). The POU domain from the ternary complex is shown in red, the Oct1/PORE complex in blue. Notice that the POU_H domain superimposes well only for the monomer bound to the upper (octamer-like) half of the PORE motif. The other POU_H shows an offset and is moved away from the DNA.

(**B**) EMSA results with the $PORE^D$ oligonucleotide, the POU domain of Oct1, full-length OBF1 protein, and a 46 amino acid containing OBF1 peptide (ΔOBF1). The binding of the full-length coactivator (~30kD) and the OBF1 peptide (~5kD) to the Oct1/PORE dimeric complex (P1/P1) result in the formation of differently migrating ternary complexes. The absence of a complex with intermediate mobility between these two ternary complexes indicates that only one OBF1 molecule can bind to the Oct1/PORE dimer (see lane 3).

(**C**) The Oct1/PORE complex could contain two coactivator binding sites at first glance. The figure shows the Oct1/PORE crystal structure with the superimposed coactivator peptides in red. Due to the different geometric parameters of the PORE DNA half-sites, the lower site is not capable of binding OBF1 (crossed out; see text).


A
H
S
S
H
B
Oct-1 + + + +
ΔOBF-1 - + + -
OBF-1 - - + +
P1/P1/OBF-1
P1/P1/ΔOBF-1
P1/P1
1 2 3 4
C
octamer half-site
OBF-1/A
H
S
S
H
OBF-1/B
nonoctamer half-site
POU1/OBF1/PORE model

3.3.4 OBF1 rescues mutations that impair POU1 dimerization

After having seen the effect POU1 mutants have on binding the octamer motif (figure 3.3), I examined their effect on dimerization using a PORE variant, $PORE^D$, that only binds POU dimers but not monomers (referred to as O^{-1} in Botquin *et al.*, 1998; Tomilin *et al.*, 2000). While the unmutated POU1 domain formed a dimer on $PORE^D$, confirming previous data, none of the six mutants retained this capability (figure 3.5A), as predicted by the structural analysis of the POU_S-POU_H dimer interface in the POU1-PORE complex (Reményi *et al.*, 2001).

The OBF1 peptide was mixed with each of the seven Oct1 derivatives to examine its effect on dimerization. A strong heterotrimer was formed when OBF1 was added to unmutated POU1 and $PORE^D$ (figure 3.5A lane 2). This is in contrast to a weak OBF1 induced ternary complex formation with POU1 on Igκ (figure 3.3 lane 2). This finding supports the notion that the POU1 dimer is a better substrate for interaction with OBF1 than the monomer (Tomilin *et al.*, 2000). Strikingly, each POU1 mutant that failed to dimerize in the absence of OBF1, formed ternary complexes on the $PORE^D$ in the presence of OBF1 (figure 3.5A, even lanes). The complexes formed with m3, m11, m12 and m13 POU1 mutants were as strong as those with unmutated POU1; the m1 and m4 mutants showed about six- and four-fold reduced capability for heterotrimer formation, respectively. Similar results were obtained when the original PORE motif was used as a probe (figure 3.5B). These results demonstrate that OBF1 rescues the detrimental effect of all POU1 mutations tested on binding to $PORE^D$. Similar to the POU1/OBF1 (PO) complex formed by the Oct1 derivatives, the POU1/POU1/OBF1 (PPO) complex showed relatively little variation in binding intensity. This is in striking contrast to the high variability of POU1 monomer binding intensities.

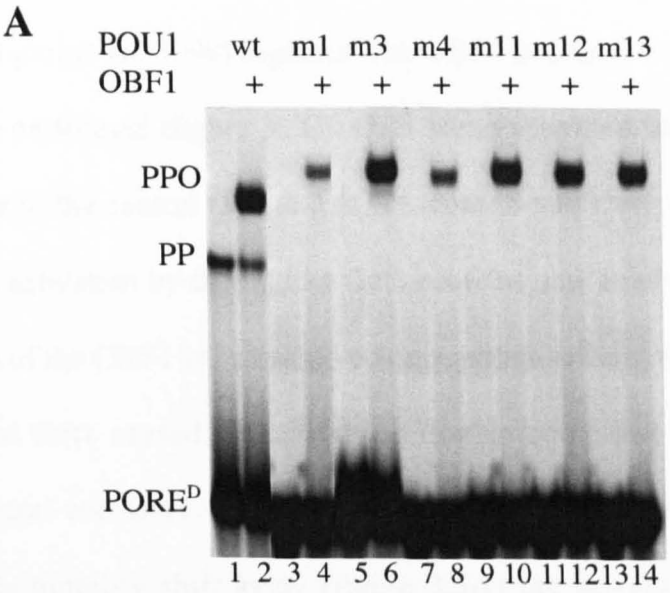
Figure 3.5

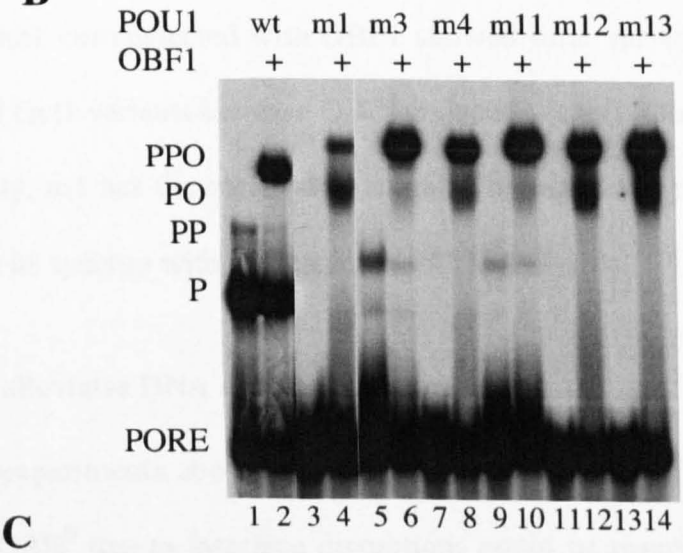
Dimer binding activities of POU1 and its derivatives under the influence of OBF1.

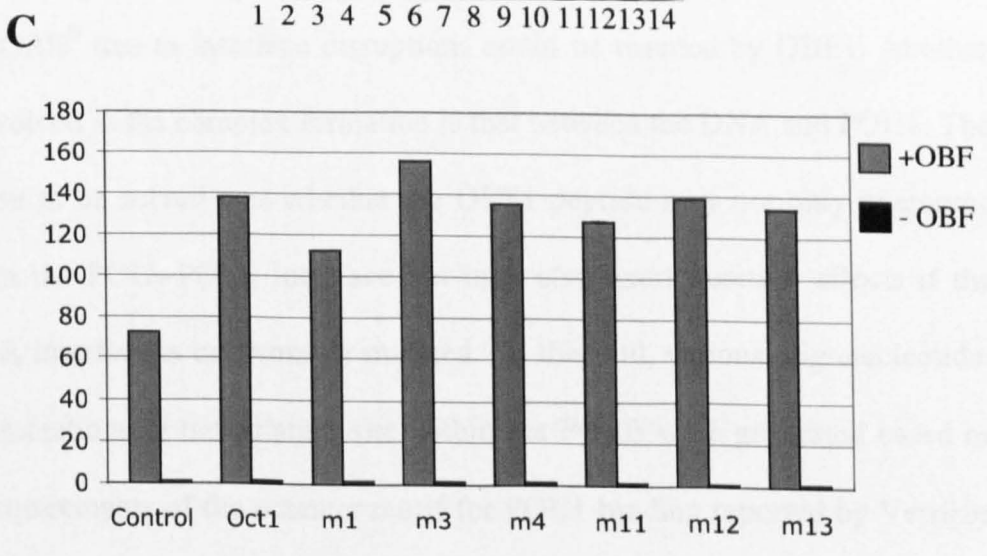
(**A**, **B**) EMSA showing effect of mutants and OBF1 on dimer binding on $PORE^D$ (**A**) and PORE (**B**). Setup as in figure 3.3A. PPO: POU1/POU1/OBF1/DNA complex; PO: POU1/ OBF1/DNA complex; PP: POU1/POU1/DNA complex; P: POU1/DNA complex. The OBF1 containing complex on the $PORE^D$ consists of two POU1 molecules and one OBF1 molecule (Tomilin *et al.*, 2000; Reményi *et al.*, 2001).

The unmutated POU1 protein exhibits a slightly higher mobility than the mutant proteins, which have an additional 3kD histidine tag. The tag slows down the migration in the gel, but does not affect interaction with the DNA.

(**C**) Comparison of Oct1 variants' transactivities in transient transfection experiments. 293 cells were transfected with $PORE^D$ luciferase reporter plasmid, wild-type or mutant Oct1 expression vectors (none in the control) +/- OBF1 expression vector (X axis). Y axis: activation of transcription, expressed as relative luciferase activities. This result was repeated 3 times.


A
POU1 wt m1 m3 m4 m11 m12 m13
OBF1 + + + + + + +
PPO
PP
PORE^D
1 2 3 4 5 6 7 8 9 10 11 12 13 14


B
POU1 wt m1 m3 m4 m11 m12 m13
OBF1 + + + + + + +
PPO
PO
PP
P
PORE
1 2 3 4 5 6 7 8 9 10 11 12 13 14


C
180
160
140
120
100
80
60
40
20
0
+OBF
- OBF
Control
Oct1
m1
m3
m4
m11
m12
m13

To address the question of whether the Oct1 mutants could also be rescued *in vivo,* cotransfection experiments of a $PORE^D$ luciferase reporter plasmid (referred to as $6xO^{-1}$ in Botquin *et al.*, 1998) together with OBF1 and Oct1 expression vectors into 293 cells were performed (figure 3.5C). Oct1 alone activated luciferase activity very little compared to the control (figure 3.5C cf. control and Oct1 -OBF and Tomilin *et al.*, 2000) and activation by the mutant Oct1 proteins was even smaller (-OBF bars). Upon addition of the OBF1 expression vector, reporter activity was highly stimulated. The unmutated Oct1 caused the activity to double compared to OBF1 transfected alone (cf. control and Oct1 +OBF). In agreement with the results obtained in the electrophoretic mobility shift assay (figure 3.5A) the activity of the mutated and unmutated Oct1 cotransfected with OBF1 showed little variation in transactivation intensity. All Oct1 variants increase OBF1-induced activity about two-fold. As in the gel shift assay, m1 has the smallest potential. The high background of OBF1 alone could be due its synergy with endogenous Oct1 in 293 cells.

3.3.5 OBF1 alleviates DNA sequence requirements

The experiments above showed that mutations that abolish dimerization of POU1 on $PORE^D$ due to interface disruptions could be rescued by OBF1. Another interface involved in the complex formation is that between the DNA and POU1. The next question to be solved was whether the OBF1 peptide may not only overcome mutations in the POU_S-POU_H interface but may also exert positive effects if the protein-DNA interface is unfavorably mutated. To this end, various oligonucleotides containing mutations in the octamer site within the PORE were generated based on sequence requirements of the octamer motif for POU1 binding reported by Verrijzer

et al. (1992). Then the formation of POU1 monomers, dimers and OBF1-induced complexes on these PORE derivatives was examined.

$PORE^D$ is one of these unfavorable mutations. Here the T at the second position of the octamer motif is mutated into a G (T2G). The POU1 monomer cannot bind, whereas the POU1 dimer can; a heterotrimer with OBF1 binds even more efficiently (figure 3.5A). Further DNA mutants are shown in figure 3.6A and B. The A1C PORE mutant generally impairs POU1 binding in the absence of OBF1. Upon addition of OBF1, however, a heterotrimer is formed, which binds as strongly to the mutant PORE as it does to the wild-type PORE sequence. Similar data are obtained for the PORE mutants A6G and A7C (referred to as O^{-4} in Botquin *et al.*, 1998; figures 6A+B), albeit some weak monomer binding is obtained even in the absence of OBF1. In addition, on the $PORE^M$ – a sequence highly related to $PORE^D$, where the octamer is intact but the T 5bp upstream of the octamer within the PORE is mutated into a G – no heterotrimer, but a weak heterodimer was formed (figure 3.6B). Furthermore, on PORE A5T, a heterotrimer bound significantly weaker than on the wild-type PORE (figure 3.6A lanes 7 and 8). This supports previous data (Gstaiger *et al.*, 1996; Cepek *et al.*, 1996; Chasman *et al.*, 1999), showing that the A5T mutation generally reduces OBF1 binding. None of these octamer-related motifs were identified as target sequences by Verrijzer *et al.* (1992), suggesting that OBF1 dramatically alleviates DNA sequence requirements of POU1 on PORE-derived elements.

3.3.6 OBF1/POU1 dimer tolerates changes in PORE binding site separation

The PORE P+1 sequence contains an insertion of one nucleotide between the two halves of the element. The Oct4 monomer binds to P+1 as well as to the wild-type PORE, whereas Oct4 dimers cannot form (Botquin *et al.*, 1998). Here the finding

Figure 3.6

Sequence requirement alleviation mediated by OBF1

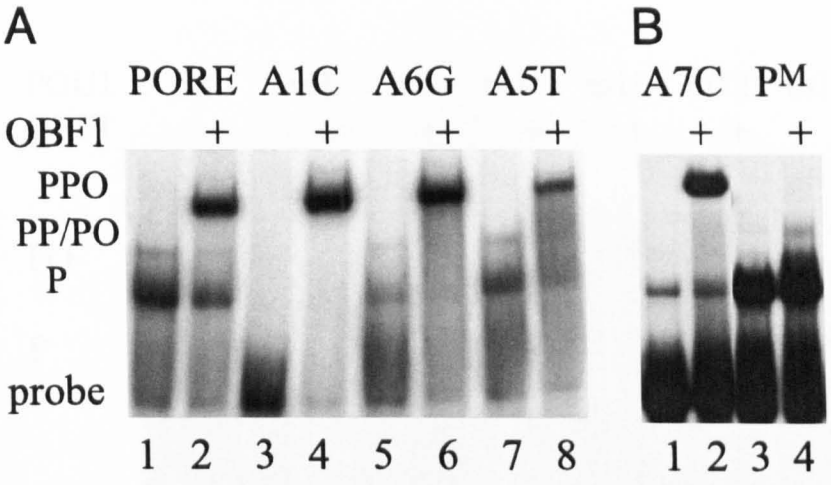
(**A, B**) EMSA showing OBF1 rescuing POU1 binding on mutant POREs. Odd lanes: Binding pattern of POU1 to the PORE and derivatives thereof. Even lanes: Binding pattern of OBF1 plus POU1 to the PORE and derivatives thereof. The table shows the specific point-mutated PORE derivatives and their abilities to bind the POU1 monomer (P), POU1 homodimer (PP), and the POU1/POU1/OBF1 complex (PPO).

(**C**) EMSA comparing POU1 with (even lanes) and without OBF1 (odd lanes) on Igκ, PORE and derivatives thereof used to demonstrate differences and similarities in their mobilities through the gel.

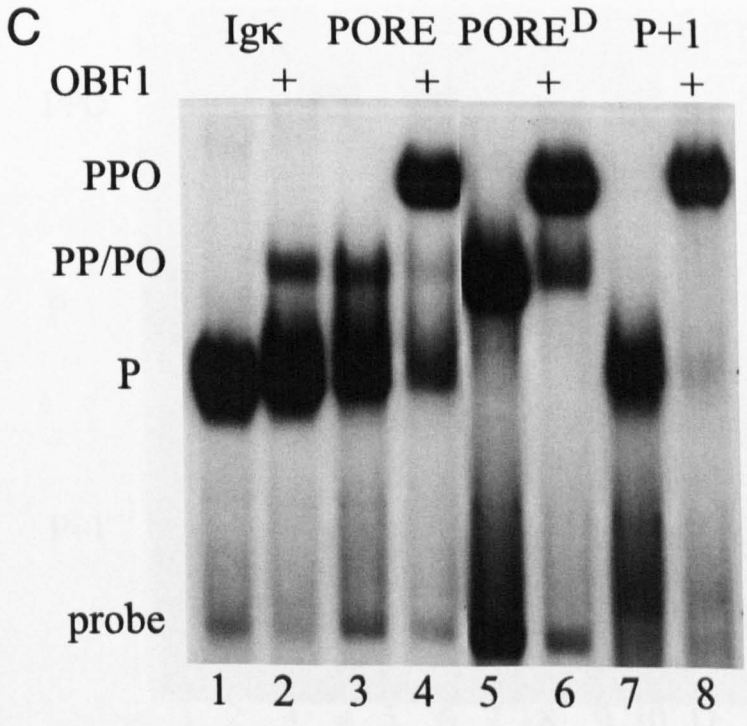
(**D, E**) EMSA with oligonucleotide $PORE^{M}$ (**D**) and P+1 (**E**). Odd lanes: Binding pattern of POU1 and its mutants, respectively. Even lanes: Binding pattern of OBF1 plus POU1 and its mutants, respectively.

The unmutated POU1 protein exhibits a slightly higher mobility than the mutant proteins, which have an additional 3kD histidine tag. The tag slows down migration in the gel, but does not affect interaction with the DNA.

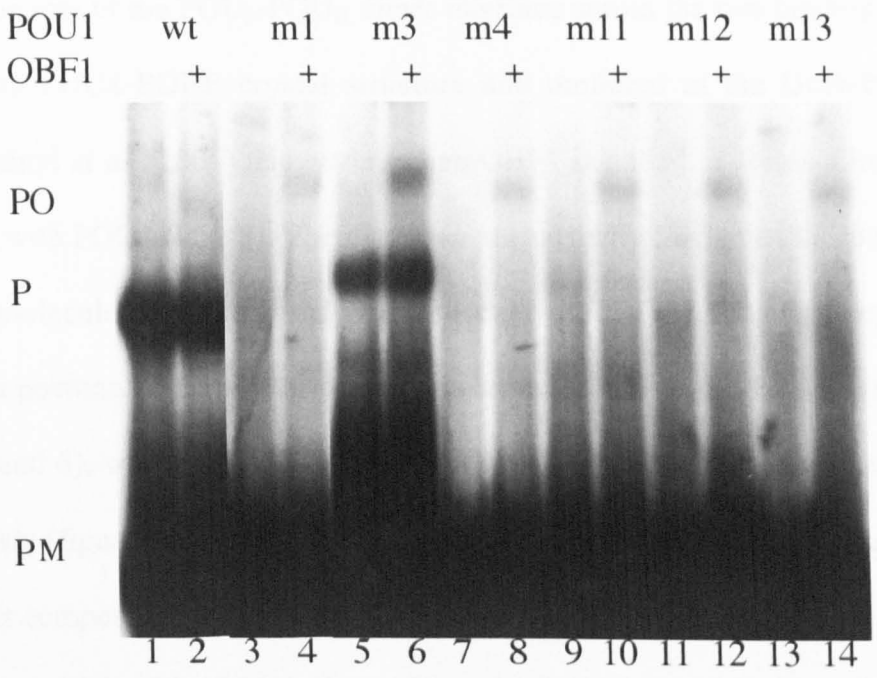
Abbreviations: P: POU1/DNA complex; PP: POU1/POU1/DNA complex; PO: POU1/OBF1/DNA complex; PPO: POU1/POU1/OBF1/DNA complex; P+1: PORE with one nucleotide inserted between the two half-sites.



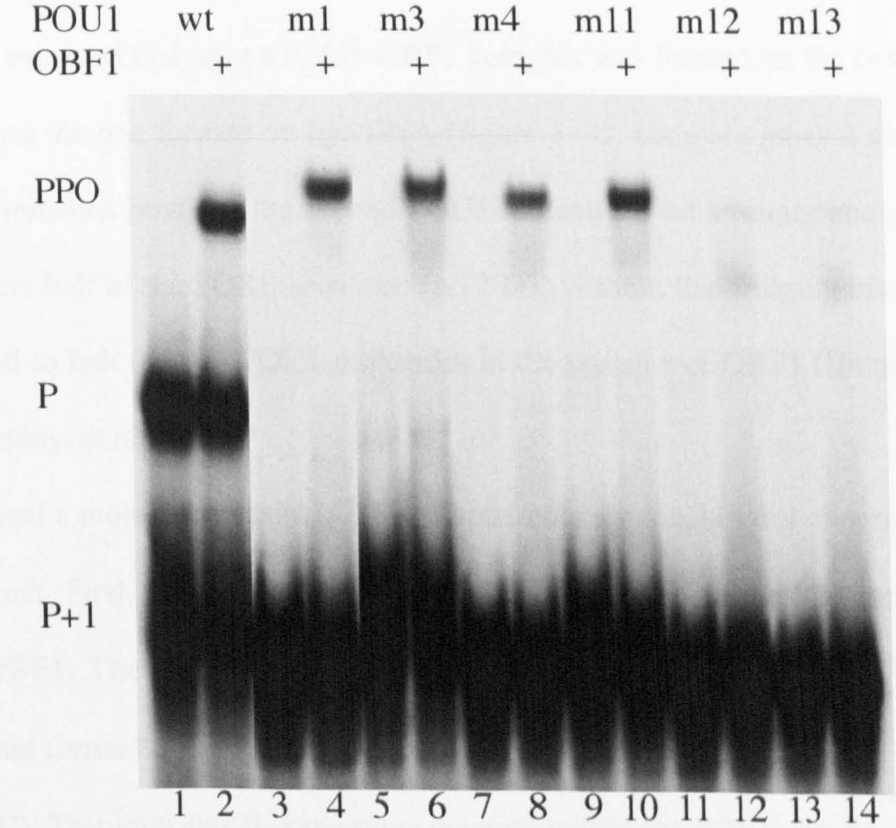
	-7 1 8	P	PP	PPO
PORE	ATTTGAAATGCAAAT	++	+	++
A1C	-------C-------	-	-	++
A6G	----------G--	+	-	++
A5T	----------T---	+	+	+
A7C	----------C-	+	-	++
PORE^{M}	--G----------	++	-	-
PORE^{D}	--------G------	-	++	++



D


POU1 wt m1 m3 m4 m11 m12 m13
OBF1 + + + + + + +
PO
P
PM
1 2 3 4 5 6 7 8 9 10 11 12 13 14

E


POU1 wt m1 m3 m4 m11 m12 m13
OBF1 + + + + + + +
PPO
P
P+1
1 2 3 4 5 6 7 8 9 10 11 12 13 14

is extended to Oct1, showing that its POU domain only forms a monomer and not a dimer on P+1 (figure 3.6C, lane 7). For both, POU1 and Oct4, these data are explained by the loss of the POU_S-POU_H dimer interface across the two binding sites observed in the POU1-PORE crystal structure and modeled in the Oct4-PORE complex (Reményi *et al.*, 2001). However, when OBF1 is added, it forms a higher-order complex with POU1 (lane 8). There are two reasons why I believe this complex contains two molecules of POU1 and one molecule of OBF1. First, the complex migrates to the position of the higher-order complex with PORE and $PORE^D$ (figure 3.6C, lanes 4 and 6), which was described as a POU1/POU1/OBF1 heterotrimer on DNA, previously (figure 3.3; Tomilin *et al.*, 2000; Reményi *et al.*, 2001). Second, the complex that is composed of only one POU1 molecule and one OBF1 migrates faster (lane 2), to the position of the POU1 dimer (figure 3.6C, lanes 3 and 5). Therefore, on P+1, OBF1 promotes binding of two POU1 molecules. This result was surprising, because I had expected that only a POU1-OBF1 complex was formed on the octamer motif, similar to the one formed on Igκ DNA (figure 3.6C, compare lanes 2 and 7). The question remains: how can the second POU1 molecule bind strongly enough to the non-octamer half of the PORE sequence (ATTTG) without the bridging interface that is required to link the two POU1 molecules in the absence of OBF1 (Botquin *et al.*, 1998; Reményi *et al.*, 2001)?

To reveal a molecular rationale for this apparent paradox, several experiments were carried out. First, binding of POU1 to $PORE^M$ was tested in the presence and absence of OBF1. The position of the POU1-OBF1 complex on the $PORE^M$ was identical to that formed on Igκ but the intensity even weaker than for Igκ (cf. figure 3.6B and 3.6C). This indicates that the same octamer within the PORE can be a good or a poor substrate for the POU1-OBF1 complex, depending on whether the second

POU1 is able to bind or not. The m4 mutant protein bound weakly to Igκ (figure 3.3, lane 7), whereas on $PORE^M$ no binding was observed (figure 3.6D, lane 7). Therefore, I considered it unlikely that m4 would bind to P+1 comparable to the other POU1 variants, if a mere structural distortion were responsible for the POU1 dimer-OBF1 complex formation on P+1. In contrast, if the two POU1 molecules are arranged parallel to the DNA axis in the presence of OBF1, and this new arrangement represents a better substrate for OBF1, m4 may bind well on P+1 in the presence of OBF1. The seven POU1-variant proteins were therefore tested for binding on P+1 in the absence and presence of OBF1 (figure 3.6E). None of the mutants could form a monomer on P+1 (figure 3.6E, odd lanes), which is in contrast to monomer binding on the Igκ site (figure 3.3). This was not due to the nucleotide insertion, as they were also unable to form on the PORE, except for m3 and m11, which bound very weakly (figure 3.5B). Only when OBF1 was added, did complexes form with the unmutated POU1, m1, m3, m4 and m11, which, according to their position on the gel, presumably contain two molecules of POU1 and one OBF1 (figure 3.6E, even lanes; cf. figure 3.6C). Since the complexes of all four mutant proteins are similar in binding intensity to P+1, binding of m4 due to structural distortion seems to be an unlikely explanation. I consider it more likely that the POU subdomains are positioned so that the linkers between them are arranged parallel to the DNA axis (also see discussion section 3.4).

Only m12 and m13 binding to P+1 could not be rescued by OBF1 (figure 3.6E, lanes 11-14). These mutant proteins have an aspartic and glutamic acid instead of a serine at position 107, respectively. While OBF1 helped tolerate these substitutions on the PORE, it appears to have reached its limit to clamp proteins with these bulky and negatively-charged side chains to P+1.

3.3.7 OBF1 stabilizes POU1 dimer on DNA by reducing its dissociation rate

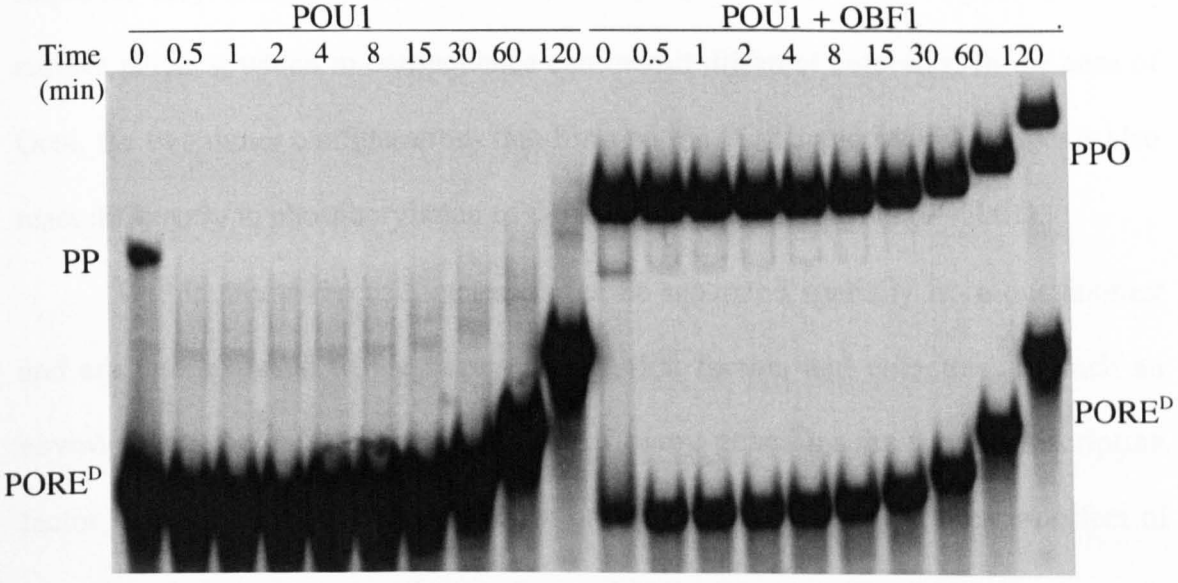
The data presented so far indicate that OBF1 has a stabilizing effect on the POU1 dimer on the PORE. Off-rate experiments were performed to further assess this effect. To this end, a binding mix was prepared into which a 500-fold excess of unlabelled oligonucleotide was added. The reaction was loaded onto a gel at regular time intervals. An increasingly weaker intensity of the bands with time reflects proteins dissociating from the DNA and being adsorbed by the excess of unlabeled oligonucleotide, thus becoming undetectable.

The POU1 dimer dissociates from the $PORE^D$ within 30 seconds of binding to it in the absence of OBF1 (figure 3.7, left half). This indicates a high off-rate and kinetic instability of the protein-DNA complex. The heterotrimer containing the OBF1 peptide forms a much higher intensity band than the POU1 dimer in the absence of OBF1 (compare 0 minute lanes in right and left halves). Furthermore, the coactivator dramatically reduces the off-rate of the POU1 dimer (right half). Even two hours after addition of the unlabeled oligonucleotide, about 50% of the complex is still bound to the labeled probe, indicating that it has not dissociated from the DNA during that time. This reflects a greater than 200-fold reduction in the off-rate with OBF1. This is in striking contrast to the lack of stabilization of the Oct1 monomer by OBF1 in similar off-rate experiments (Strubin *et al.*, 1995), again indicating that the PORE-type Oct1 dimer is a preferred substrate of the coactivator compared to the POU1 monomer.

Figure 3.7

Off-rate EMSA of POU1 dimer vs. OBF1 mediated heterotrimer

Showing the stability of the POU1 dimer and the POU1/POU1/OBF1 complex on DNA. A binding reaction is prepared into which a 500-fold excess of unlabeled oligonucleotide is added after the proteins have been allowed to bind the labeled probe. Aliquots of the reaction mixture are loaded onto a gel at regular time intervals between 30 seconds and 120 minutes after addition of the unlabeled oligonucleotide. Dissociating proteins become undetectable on the gel as they re-associate to the unlabeled oligonucleotide.


POU1
POU1 + OBF1
Time (min)
0 0.5 1 2 4 8 15 30 60 120 0 0.5 1 2 4 8 15 30 60 120
PPO
PP
PORE^D
PORE^D

3.4 Discussion

3.4.1 Osteopontin in B-cells: gene regulation and function

POU factors exert a high level of flexibility in regulating gene expression, attributed in part to their ability to bind DNA as monomers, homo- and heterodimers. In addition, these dimers can adopt different configurations depending on the DNA sequence they bind to (Tomilin *et al.*, 2000). In doing so, they expose different surface patches, which in consequence can recruit different cofactors. In the case of Oct4, the two dimer configurations that form on the PORE and MORE elements also react differently to phosphorylation of the POU factor (Reményi *et al.*, 2001).

Within any given cell, genes cannot be separated spatially from one another and are thus exposed to the same transcription factors and cofactors. In such an environment, the differential regulation of many genes by the same transcription factor must occur. This may be at least partly attributed to the differential effect of phosphorylation on transcription factor binding to specific DNA elements in conjunction with the selective recruitment of cofactors. It was previously reported that the heavy chain of the immunoglobulins is regulated by the MORE-type dimer in lymphoid cells (Tomilin *et al.*, 2000) and in this study it is established that osteopontin is a target gene of the PORE-like dimer and OBF1 in the same cells. Thus, we found an environment conducive to various degrees of differential gene regulation.

Secreted OPN stimulates B-cells to produce immunoglobulins and, together with an unidentified 14kD peptide, to proliferate (for review see Weber and Cantor, 1996). Still, the precise function of osteopontin in B-cells remains elusive. In the absence of OBF1, osteopontin is expressed at very low levels (figure 3.1A) and might consequently not be able to stimulate B-cells to proliferate, leading to reduced

numbers of mature B-cells in OBF1-deficient mice. This reduction of B-cell number in OBF1 deficient mice has been reported by Kim *et al.* (1996), Nielsen *et al.* (1996) and Schubart *et al.* (1996). Thus, osteopontin, along with the 14kDa peptide, might have an autocrine effect on B-cells.

3.4.2 OBF1 alleviates DNA sequence requirements for POU dimerization and stabilizes the dimer on the DNA

Verrijzer *et al.* (1992) analyzed the binding specificity of POU1 by screening a library of randomly synthesized oligonucleotides with POU1 for high affinity targets. Consequently, a sequence similar to the octamer motif was defined as the Oct-binding consensus. Based on this consensus, I generated various oligonucleotides with unfavorable mutations within the PORE and tested them for monomer, dimer and OBF1-induced complex formation. I found that even though monomer and dimer binding were often impaired, OBF1 would discount these unfavorable sequences and clamp the POU1 dimer onto the DNA (figure 3.6A). The lack of an OBF1-induced complex on the $PORE^{M}$ and the presence of a weak one on A5T proved that the cofactor does not bind indiscriminately. Nonetheless, OBF1 alleviates sequence requirements substantially. This implies that a weak binding site might only be activated by the Oct1 dimer in the presence of OBF1 whereas a strong PORE-type binding site may not require a cofactor to activate the target gene.

The results obtained with P+1 – an element in which the two Oct-factor binding sites have been moved apart by the insertion of one nucleotide – indicate that OBF1-induced dimerization tolerates changes in the separation of the two half binding sites on PORE (figure 3.6C). Botquin *et al.* (1998) showed that on P+1 dimerization cannot occur, and proposed that this was a consequence of the interface having been lost between the two protein molecules. OBF1 overcomes these binding

difficulties and mediates heterotrimer binding with the POU1 dimer to the oligonucleotide. There are two possible explanations to account for this ability of OBF1: (i) OBF1 may clamp the two molecules onto the DNA at a less favorable binding site so that the POU factors are bound like on the PORE, relative to each other, ignoring the phasing mutation. (ii) The POU1 molecules bind to the same nucleotides on P+1 as on the PORE but protein-protein interaction is not required, because OBF1 tethers the two molecules to the DNA. This has to be such a strong interaction that the POU1 molecules do not have to support each other to remain bound to the DNA.

I favor the second explanation partly because of the off-rate results (figure 3.7), which revealed that the OBF1 peptide has a strong stabilizing effect on the POU1 dimer. This solid stabilizing potential of OBF1 had not been detected on the POU1 monomer. Babb *et al.* (1997) showed a slight OBF1 stabilizing effect on the POU1 monomer by DNase I footprint assays. On the other hand, Luo and Roeder (1995) could not demonstrate an effect of OBF1 on POU1 stability by titration studies. Furthermore, Strubin *et al.* (1995) showed that the dissociation rate of the Oct1 monomer from the octamer binding site was not influenced by the presence of OBF1, even though the cofactor exhibits a clamping effect by securing the two separate POU_S and POU_H domains onto the DNA, as shown by EMSA.

In addition to the PORE-like sequence, the flanking nucleotides also influence the binding ability of the POU1 dimer and OBF1. POU1 does not bind $PORE^D$ as a monomer, but if the two nucleotides 3' of the element are mutated from GG to TA it does (figure 3.8A, PD->Igκ). The fact that none of the 56 target sites selected for POU1 binding by Verrijzer *et al.* (1992) contained two Gs following the octamer motif may reflect how POU proteins need to interact as monomers and dimers with

Figure 3.8

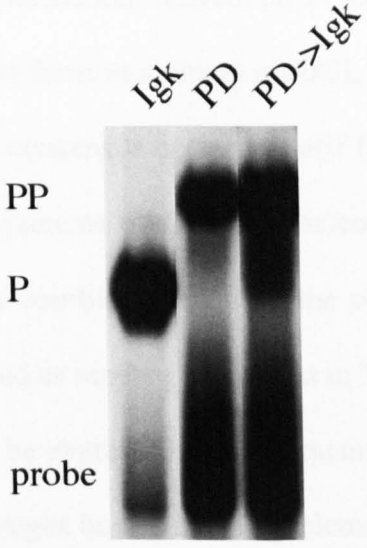
Influence of octamer flanking sequence

(**A**) EMSA comparing POU1 binding to Igκ, $PORE^D$ (PD) and $PORE^D$->Igκ (PD->Igk), a derivative of the PORE with Igκ-like 3'flanking nucleotides.

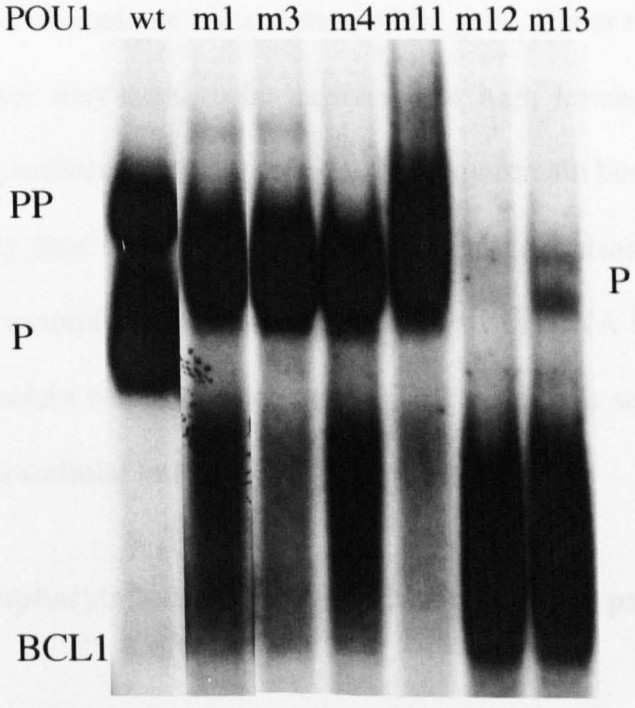
(**B**) EMSA of POU1 and mutants thereof on BCL1, a MORE, which occurs in numerous human and mouse V_H promoters (Tomilin *et al.*, 2000). The unmutated POU1 protein exhibits a slightly higher mobility than the mutant proteins, which have an additional 3kD histidine tag. The tag slows down migration in the gel, but does not affect interaction with the DNA.

(**C**) Sequences of oligonucleotides used to investigated the influence of 3' flanking region of the octamer. The octamer (in Igκ), PORE (in $PORE^D$, $PORE^D$->Igκ) and MORE (in BCL1) are underlined. The 2bp 5' flanking sequence analyzed is written in capitals, mutated nucleotides are highlighted in bold.

A



B



C

Igk:	agggtatgcaaatTAtta
PD:	atttgaaaggcaaatGGaaa
PD->Igk:	atttgaaaggcaaat**TA**aaa
BCL1:	atgaatatgcaaatCAggt

the PORE. This supports the notion that the two Gs following the PORE are disadvantageous for monomer formation. Moreover, POU1 and its mutants m1, m3, m4 and m11 (cf. figure 3.2) can form monomers on BCL1 (Tomilin *et al.*, 2000 and figure 3.8B), which contains a consensus octamer motif followed by the nucleotides CA. For octamer flanking sequences of all oligonucleotides concerned see figure 3.8C. Indeed, the TA and CA combinations 3' of the octamer motif, which allow monomer binding, were obtained as binding sequences in Verrijzer's experiment.

Genes that should only be expressed in small quantities or when high levels of POU1 are present in the cell might be regulated by elements to which OBF1 cannot clamp the POU1 dimer. In such a case, the dimer can dissociate from the regulatory element rapidly and thus discontinue transcription of the gene. Genes regulated by the OBF1-Oct1 heterotrimer may need to be expressed at high levels once they are activated. This can be provided for, since the complex may remain bound to DNA for a longer time and may thus allow multiple rounds of transcription to occur. The osteopontin gene, for example, may require high levels of mRNA to be generated continuously, as the protein concentration is decreasing when it is secreted from the confined cell to the extracellular milieu.

3.4.3 Mimicking phosphorylation at the PORE dimer interface prohibits POU1-DNA interaction

One of the most frequently used mechanisms to regulate the activity of transcription factors in response to different extra- and intracellular signals is phosphorylation and dephosphorylation (Whitmarsh and Davis, 2000). Activities of several members of the POU factor family are controlled by this mechanism. Oct1 is hyperphosphorylated as cells enter mitosis. This correlates with strongly reduced Oct1 binding to the octamer site and a concomitant inhibition of transcription. Phosphorylation of Oct1 is

rapidly reversed as cells exit mitosis and enter the G1 phase of the cell cycle (Roberts *et al.*, 1991). It was shown that mitosis-specific phosphorylation of S107 in the homeodomain of Oct1 was sufficient to inhibit the DNA-binding ability of this factor (Segil *et al.*, 1991). This serine is situated in a KRTSIE motif, which is a potential site for phosphorylation by a cAMP- or cGMP-dependent protein kinase. Another member of the POU factor family, Pit1, was shown to become phosphorylated at two distinct sites in the homeodomain of the protein in pituitary cells in response to phorbol esters and cAMP. One of these sites is T107, which plays a similar role to S107 in Oct1 (Kapiloff *et al.*, 1991).

The mutations in m12 (S107D) and m13 (S107E) imitate phosphoserines (Maciejewski *et al.*, 1995). In agreement with previous observations, these mutants are unable to bind POU binding sites as both monomers and dimers (figure 3.5). The bulky and negatively-charged aspartic and glutamic acids get close to the negatively-charged DNA and inhibit contact between the amino acid and the phosphate backbone. This implies that the S107-phosphorylated Oct1 cannot bind the elements it could in the unphosphorylated state.

OBF1 overcomes the steric clash of the phosphoserine mimic with the DNA. When the cofactor is added to the binding reaction it mildly rescues the phosphorylation-mimicked POU1 mutant monomer on Igκ. On the PORE and PORE^{D} binding is more enhanced. This difference correlates with the fact that OBF1 stabilizes the dimer more than the monomer (figure 3.7 and Luo and Roeder, 1995; Strubin *et al.*, 1995). On P+1, though, the phosphorylation-mimicked POU1 dimer cannot be rescued by OBF1. Thus, in the presence of OBF1 phosphorylation of S107 could inhibit transcription of a gene regulated by a P+1-type element but not one regulated by a PORE-type sequence. By this distinction, genes could be differentially

regulated by the same transcription factor dimer depending on the presence or absence of a coactivator, the phosphorylation state of the transcription factor, and the DNA sequence of the regulatory element.

There are multiple events that trigger kinases and phosphatases, changing the expression patterns of various genes, such as cAMP elevation, membrane depolarization/calcium influx, and growth factor reception. More specifically, cAMP levels play an important role in the costimulation of lymphoid cells, which is critical for an appropriate immune response. In T-cells, elevation of cAMP levels is inhibitory. CD28 costimulation induces expression of a cAMP phosphodiesterase leading to reduced cAMP levels and thus stimulation of T-cells (for review see Frauwirth and Thompson, 2002). Similarly, cAMP could result in phosphorylated Oct1 in B-cells. Due to its phosphorylation, the transcription factor might not be able to activate all genes required to stimulate the cell, e.g., *OPN*. Upon costimulation mediated by a receptor resulting in a reduction of cAMP, Oct1 could be dephosphorylated and regulate a wider variety of genes. Incidentally, serine 107, which is phosphorylated *in vivo*, is situated within a KRTSIE motif, which is a potential site for phosphorylation by a cAMP-dependent protein kinase.

3.4.4 Novel POU1 dimer configuration upon interaction with OBF1

The data in this chapter show that OBF1 compensates for PORE interface mutations and that the cofactor stabilizes the POU1 dimer, locking it onto the DNA. Based on my observations, I speculate that OBF1 may induce the Oct1 dimer to adopt yet a different arrangement than proposed for the PORE- and MORE-type binding (Botquin *et al.*, 1998; Tomilin *et al.*, 2000; Reményi *et al.*, 2001). If the dimer binds as proposed for the PORE (figure 3.9B), why does OBF1 have such a strong effect on two POU1 molecules clamping only one of them to the DNA? How does this help the

Figure 3.9

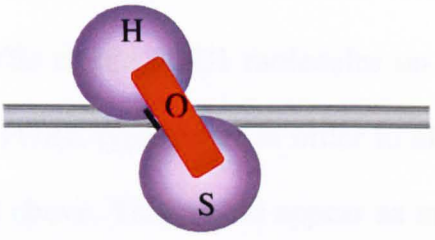
Model of POU1-OBF1 complexes bound to specific DNA elements in different configurations.

(**A**) POU1 and OBF1 binding the octamer motif according to Chasman *et al.* (1999).

(**B**) POU1 dimer configuration as proposed by Botquin *et al.* (1998) with OBF1 binding to the POU1 molecule at the octamer half-site as proposed by Reményi *et al.* (2001).

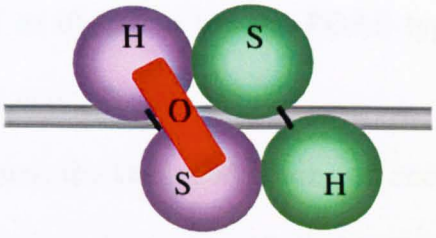
(**C**) POU1 dimer as in (B), but with a different linker connectivity. Here, OBF1 binds to the POU_S domain of one POU1 molecule and to the POU_H of another on the octamer half-site, with the individual subdomains arranged as in the crystal structure solved by Reményi *et al.* (2001).

A



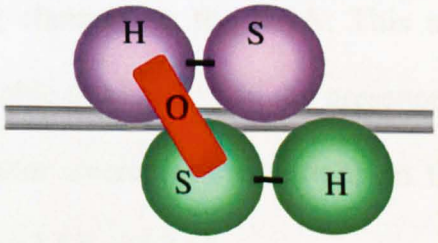
OBF1 and POU1 on Igκ

B



OBF1 and POU1 dimer:
PORE-like configuration

C



OBF1 and POU1 dimer:
Novel configuration

second molecule bind to the first and the DNA, especially if these molecules are mutated in the PORE dimer interface and bind to P+1?

I propose two possibilities for the novel configuration. One may be a scenario in which the interaction with OBF1 slightly modifies the POU1 dimer's overall configuration from that of the PORE-type dimer. In this case, the binding of OBF1 would result in a minor shuffle of the POU1 molecules on the DNA, so that they interact differently from the PORE-type dimer in order to allow for all unfavorable binding conditions discussed above. This would appear as in figure 3.9B but would show differences to the PORE-like dimer in a more detailed illustration. An alternative scenario would be that the POU1 subdomains POU_S and POU_H are bound to the DNA sequence motif as observed for the PORE-type dimer (figure 3.9B; Botquin *et al.*, 1998; Reményi *et al.*, 2001).

In the PORE-type dimer, the two subdomains of one molecule are bound to one half site, whereas I propose that with OBF1 binding, the two subdomains belonging to one molecule bind parallel to the DNA strand (figure 3.9C). Thus, OBF1 would bind to the POU_H and POU_S from two different POU1 molecules leading to the two POU1 molecules being clamped to the DNA. This model could provide a rationale as to why POU1 is able to bind P+1 in the presence of OBF1 but not in its absence, and how the cofactor overcomes mutations in the protein-protein and protein-DNA interface (figures 3.5A and 3.6).

Furthermore, the fact that OBF1 reduces the POU1 dimer off-rate from the $PORE^D$ so severely (figure 3.7) supports a novel arrangement, whose interaction between the different molecules is stronger than that of the PORE-type dimer. Moreover, this experiment does not show any PO intermediate complex forming as time progresses. A PO intermediate would have been supportive of a PORE-type

configuration, with both linkers being perpendicular to the DNA axis (figure 3.9B). Had OBF1 clamped the POU1 dimer in the PORE configuration, the POU1 molecule associated with OBF1 might have remained bound to the DNA even after the first POU1 had dissociated. In addition, the novel configuration supports the idea of the clamping effect being much more pronounced for the POU1 dimer (figure 3.7) than for the monomer (Luo and Roeder, 1995; Strubin *et al.*, 1995).

The flexible nature of the linker connecting the POU subdomains argues against resolving this issue by a structural approach, since the linker region has remained invisible in all POU/DNA crystal structures solved so far. It might still be possible to elucidate which of the two arrangements is true by determining the crystal structure of the POU1 dimer with OBF1 on the PORE. If the POU1 molecules bind like in the PORE-type dimer, the novel dimer configuration can be assumed in order to accommodate for the stabilizing impact of OBF1 on the dimer. If the subdomains cannot be superimposed on the coordinates of the PORE-type dimer, the induced-fit model might be the more likely one.

These results change our view on differential regulation and its fine-tuning via the various POU dimers in the context of coactivators. Most importantly, they highlight the limitations of a mutational *in vitro* analysis when it comes to applying the results to biological questions in cells or even living organisms. Prior to this study, one could have assumed that introducing specific dimer mutations into an endogenous POU gene might have exhibited a strong phenotype due to the total loss of dimerization *in vivo*. Now, one might anticipate a more subtle or even absent phenotype. However, only the introduction of PORE and MORE interface mutations into endogenous POU genes will provide an idea whether associated factors that facilitate dimerization *in vitro*, may also help to maintain the full regulatory program *in vivo*.

Chapter 4

DNA mediated interface swapping of POU and Sox

4.1 Abstract

Members of the POU and Sox transcription factor families exemplify partnerships established between various transcriptional regulators during early embryonic development. Although functional cooperativity between key regulator proteins is pivotal for milestone decisions in mammalian development, little is known about the underlying molecular mechanisms. In this chapter, I focus on two transcription factors, Oct4 and Sox2, as their combination on DNA is considered to direct the establishment of the first three lineages in the mammalian embryo. The data presented here validate experimental high-resolution structure determination, followed by model building. This study shows that Oct4 and Sox2 are able to dimerize on DNA in distinct conformational arrangements. The binding site characteristics of their target genes are responsible for the correct spatial alignment of the Velcro-like interaction domains on their surface. Interestingly, these surfaces frequently have redundant functions and are instrumental in recruiting various interacting protein partners.

4.2 Introduction

Transcription factors form multi-protein complexes on DNA, in order to orchestrate the correct temporo-spatial expression of developmental genes. The process leads to the establishment of functional partnerships, with the combination rather than the individual activity of each factor eliciting specific transcriptional outcomes. Members of the transcription factor families POU and Sox exemplify this functional cooperativity during early embryonic development. POU (Herr and Cleary, 1995) and Sox (Wegner, 1999) proteins selectively interact with each other via their conserved POU and HMG DNA-binding-domains. Their functional partnership has

been characterized on regulatory elements in various species, including human, mouse and the fruit fly (Dailey and Basilico, 2001). POU and Sox proteins are differentially expressed during development, and their combinations may lead to the differential expression of genes critical for cell-fate determination (Dailey and Basilico, 2001). The genes encoding the transcription factors Oct4 and Sox2 are tightly regulated during development and in embryonic cell lines (see discussion, section 4.4). Their combination is critical, as it functions to specify the first three lineages in the mammalian embryo (Nichols *et al.*, 1998; Niwa *et al.*, 2000; Avilion *et al.*, 2003). While Oct4 and Sox2 are considered to define a combinatorial code *in vivo* (Avilion *et al.*, 2003), binding of POU factors by Sox2 *in vitro* is rather indiscriminate. For example, the POU domains of several family members, including the prototype member Oct1, bind cooperatively with the HMG domain of Sox2 onto the 3' UTR of the *fibroblast growth factor 4* (*FGF4*) (Ambrosetti *et al.*, 1997). However, its activation *in vivo* is dependent on Sox2/Oct4 binding and is mediated by Oct4-specific regions external to the POU domain (Ambrosetti *et al.*, 2000). Another POU/Sox dependent element is responsible for regulating expression of the *Undifferentiated Transcription Factor 1* gene (*UTF1*) (Nishimoto *et al.*, 1999).

In this chapter I investigate the interaction of Oct1 and Oct4 with Sox2 on two different DNA binding elements to test whether a previously discovered regulation mechanism of DNA-mediated swapping of the arrangement of homodimers may also be applicable for unrelated transcription factor assemblies (Tomilin *et al.*, 2000; Reményi *et al.*, 2001). The crystal structure of the ternary Oct1/Sox2/*FGF4* element complex was solved and homology modeling tools were used to construct an Oct4/Sox2/*FGF4* as well as an Oct4/Sox2/*UTF1* structural model, as discussed in this chapter. These models and the biochemical data verifying them revealed that the

FGF4 and the *UTF1* elements mediate the assembly of distinct POU/HMG complexes leading to different quaternary arrangements by swapping protein-protein interaction surfaces of Sox2. Moreover, data in this chapter show that Sox2 uses one of these two protein interacting surfaces to assemble a ternary complex with another unrelated transcription factor on a late embryonic stage specific enhancer (Pax6/Sox2 on the *DC5* element). The findings of this chapter outline a simple mechanism for promiscuous yet highly specific assembly of transcription factors, in which the sequence of DNA enhancers governs a combinatorial use of redundant protein-protein interaction surfaces.

4.3 Results

4.3.1 Oct4 and Sox2 interact differentially on the *FGF4* and *UTF1* element

Sox2 and Oct4 interact with each other on the *FGF4* and the *UTF1* elements in order to activate expression of the corresponding genes (Yuan *et al.*, 1995; Nishimoto *et al.*, 1999). A comparative titration of Sox2 with Oct4 on the *FGF4* and *UTF1* elements in an electrophoretic mobility shift assay (EMSA) revealed, that Oct4 and Sox2 interact differently with each other on these two elements *in vitro*. A lower amount of Sox2 HMG domain was required for heterodimerization with Oct4 on *UTF1* than on *FGF4* (figure 4.1A). Since only the POU domain of Oct factors is often used for experiments to investigate binding properties of Oct factors and to solve crystal structures I investigated whether the Oct4 POU domain (POU4) behaves similarly. Indeed, it was found that POU4 is sufficient to exert this differential cooperativity and the N- and C-terminal transactivation domains are not required here (figure 4.1B).

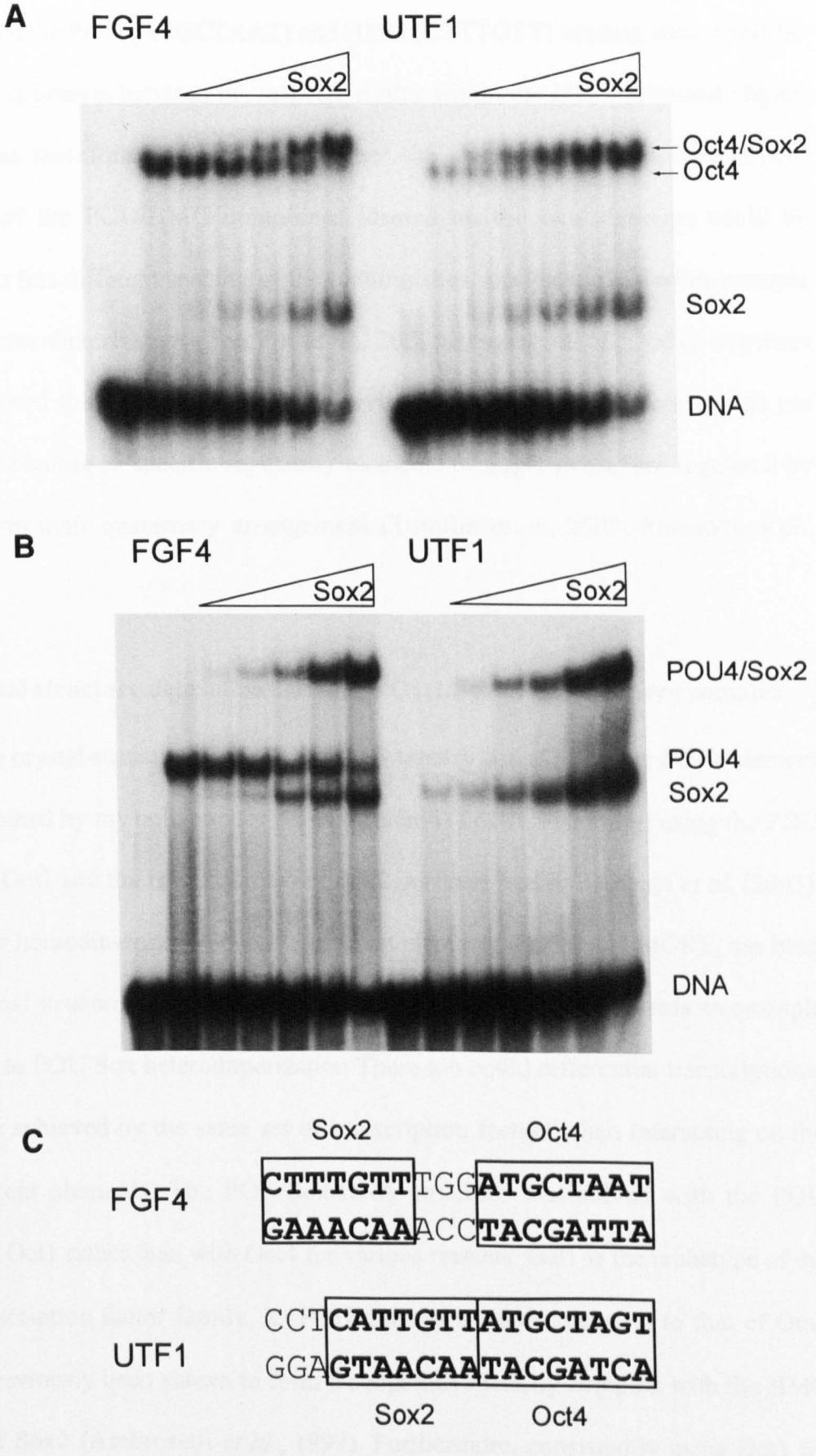
Figure 4.1

Oct4 and Sox2 interact differentially on *FGF4* and *UTF1*

(**A**) EMSA assay of Oct4 and Sox2 with radiolabeled DNA oligonucleotides of *FGF4* and *UTF1*. Lane 1: no protein; lane 2: Oct4; lane 3-7: increasing amounts of Sox2-HMG protein mixed with equal amounts of Oct4. Although less Oct4 binds alone onto *UTF1* DNA than onto *FGF4*, heterodimerization on *UTF1* is more pronounced with lower amounts of Sox2 than on *FGF4*. The significantly lower degree of Oct4/DNA interaction in the absence of Sox2 on *UTF1* compared to that of on *FGF4* is very likely due to the non-optimal octamer motif sequence for POU binding within the *UTF1* element (Nishimoto *et al.*, 1999).

(**B**) EMSA as in figure 4.1A but only with the POU domain of Oct4 (POU4).

(**C**) Differential spacing of binding sites for Oct4 and Sox2 in *FGF4* and *UTF1*.


A
FGF4
UTF1
Sox2
Sox2
Oct4/Sox2
Oct4
Sox2
DNA
B
FGF4
UTF1
Sox2
Sox2
POU4/Sox2
POU4
Sox2
DNA
C
Sox2
Oct4
FGF4
CTTTGTTTGGATGCTAAT
GAAACAAACCTACGATTA
UTF1
CCTCATTGTTATGCTAGT
GGAGTAACAATACGATCA
Sox2
Oct4

The *FGF4* element, located within the 3' UTR of the gene, contains 3 base pairs between the POU (ATGCTAAT) and HMG (CTTTGTT) binding sites, while no such spacer is present between the respective sites within the *UTF1* promoter (figure 4.1C). I was therefore interested in whether the observed different biochemical properties of the POU/HMG complexes formed on the two elements could be attributed to this different spacing of the binding sites, similar to the earlier example of POU factor dimerization (Tomilin *et al.*, 2000; Reményi *et al.*, 2001). Previous studies showed that the transcriptional activities of POU factor dimers, which are induced by binding to specific regulatory elements of target genes, are regulated by alterations in their quaternary arrangement (Tomilin *et al.*, 2000; Reményi *et al.*, 2001).

4.3.2 Crystal structure determination of the Oct1/Sox2/*FGF4* ternary complex

The crystal structure of the POU/HMG ternary complex on the *FGF4* element was determined by my collaborator Attila Reményi (EMBL Hamburg) using the POU domain of Oct1 and the HMG domain of Sox2, as described in Reményi *et al.* (2003). POU factor homodimerization on two different elements, PORE and MORE, has been characterized structurally with the POU domain of Oct1, and it represents an example analogous to POU/Sox heterodimerization There too could differential transcriptional activity be achieved by the same set of transcription factors when interacting on the two different elements. The POU/Sox/DNA structure was solved with the POU domain of Oct1 rather than with Oct4 for various reasons. Oct1 is the archetype of the POU transcription factor family, its POU domain is ~60% identical to that of Oct4 and has previously been shown to form a cooperative ternary complex with the HMG domain of Sox2 (Ambrosetti *et al.*, 1997). Furthermore, consistently using Oct1 for structural studies allows the direct comparison of POU homodimerization and

POU/Sox heterodimerization on different DNA elements. The 3'UTR of the *FGF4* gene was the first DNA element that was described to contain a composite DNA element binding Sox2 and Oct4 (Yuan *et al.*, 1995). Further biochemical work demonstrated the cooperative nature of this interaction with Oct4 as well as with Oct1 (Ambrosetti *et al.*, 1997).

The crystal structure of the Oct1/Sox2/*FGF4* ternary complex reveals a novel heterotrimeric domain arrangement, in which the centrally positioned POU specific domain (POU_S) interacts with the HMG domain of Sox2 and the POU homeodomain (POU_H) of Oct1 (figure 4.2A). The Sox2 HMG domain adopts an L-shaped structure and its N-terminal 70-residue segment folds like other structurally characterized HMG domains (Weiss, 2001).

Sox2 binds in the minor groove of the DNA and forms an HMG/DNA interaction surface that is comparable in size to that of the POU/DNA (1350 $Å^2$ and 1400 $Å^2$, respectively) (figure 4.2B). The size of an interaction surface reflects the strength of the van-der-Waals interactions involved. Even though the HMG domain is a lot smaller than the POU domain they interact with the DNA with comparable surface sizes. In contrast to the straight POU binding site of *FGF4*, the Sox2-HMG domain bends its cognate sequence towards the major groove in an angle of about 90°, similar to what is reported of other Sox factors (Weiss, 2001). Figure 4.2B depicts the sequence specific hydrogen bonds, which contribute to the protein-DNA interaction. The amino acids responsible for bending the DNA by inserting their hydrophobic side chains between two basepairs are also shown. Side chains from residues of α-helix 1 and α-helix 2 of the Sox2-HMG domain are inserted between 3-bp stacks of the recognition sequence (C^T^T^TGTT; figure 4.2B), leading to unwinding of the DNA α-helix at the Sox2 binding site. This insertion and unwinding

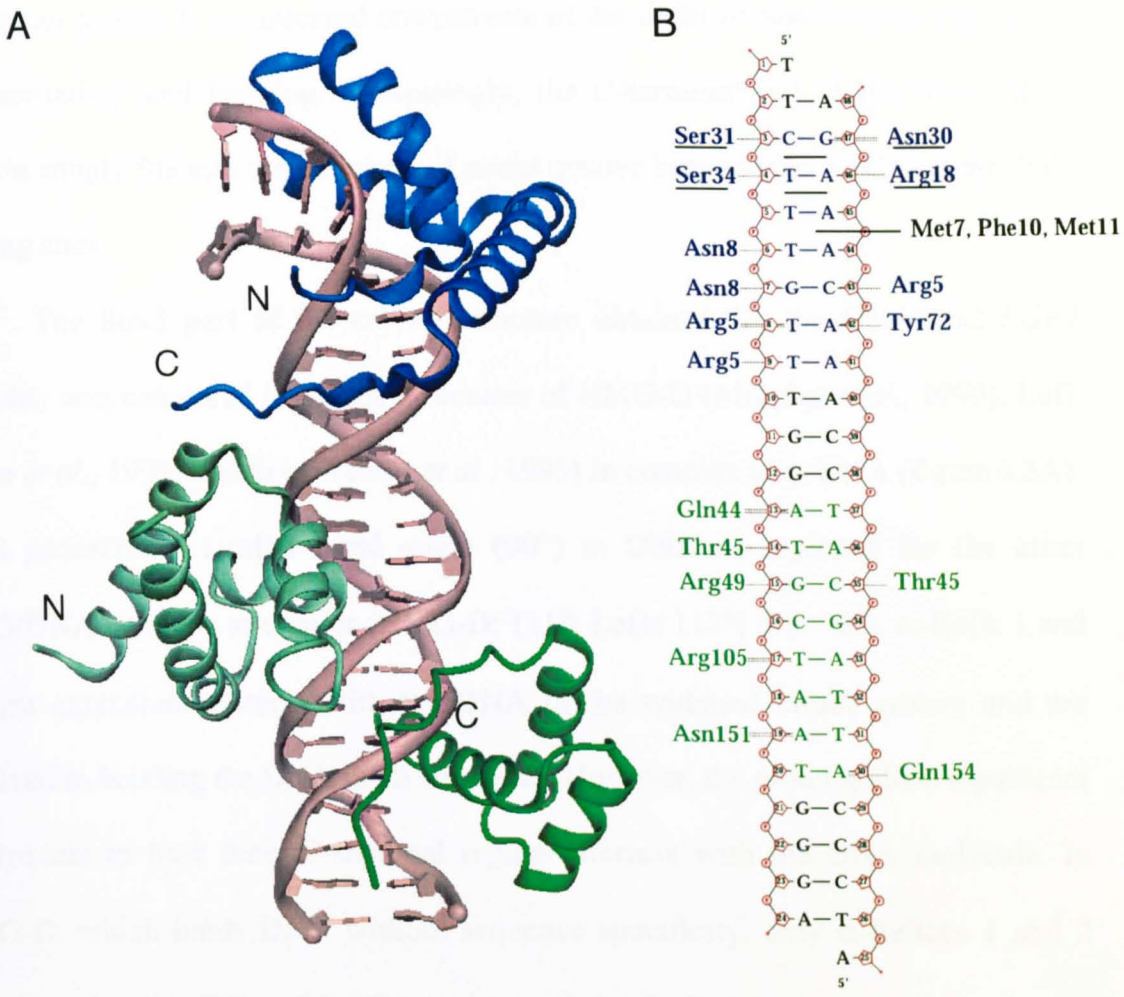
Figure 4.2

Crystal structural of the Oct1/Sox2/*FGF4* ternary complex

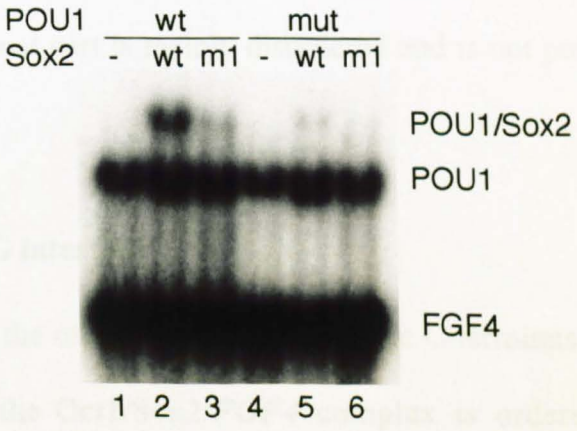
(**A**) The centrally positioned POU specific domain (POU_S) interacts with the HMG domain of Sox2. The HMG domain of Sox2 is colored in blue and the POU domain of Oct1 in green (POU_S: light green; POU_H: dark green). Some part of the POU linker region connecting POU_S and POU_H is invisible in the crystal structure.

(**B**) Protein-DNA interactions between HMG, POU1 and the *FGF4* element. The figure shows the sequence of the oligonucleotide used for crystallization. The HMG and the POU domain binding sites are colored blue and green, respectively. For simplicity, only DNA sequence-specific hydrogen bonds are shown. The hydrophobic side chains of M7, F10 and M11 are inserted between base pairs T5•A45 and T6•A44 causing a ~45° bend of the DNA axis. These and other amino acid residues that play a role in bending the DNA at three different base stack levels are either underlined or written in black.

(**C**) EMSA results of Oct1/Sox2/DNA ternary complex formation on the *FGF4* element using mutant versions of the HMG domain of Sox2 (m1) and POU domain of Oct1 (mut). The mutants were designed to interfere with the POU-HMG interface formation based on the Oct1/Sox2/*FGF4* crystal structure. The R75E mutation in Sox2 (m1) and the I21Y,D29R mutation in POU1 (mut) were introduced to reverse the charges or increase the bulkiness of important amino acids within the POU/HMG interface. wt: wild-type protein.


A
N
C
N
C
B
Ser31
Asn30
Ser34
Arg18
Met7, Phe10, Met11
Asn8
Asn8
Arg5
Arg5
Tyr72
Arg5
Gln44
Thr45
Arg49
Thr45
Arg105
Asn151
Gln154

C


POU1
wt
mut
Sox2
- wt m1 - wt m1
POU1/Sox2
POU1
FGF4
1 2 3 4 5 6

results in the minor groove becoming shallow and expanded at the Sox binding site. The minor groove is compressed downstream of the widened Sox binding site due to the unwinding and bending. Interestingly, the C-terminal tail of the Sox2-HMG domain snugly fits into the compressed minor groove between the Sox2 and the POU binding sites.

The Sox2 part of the crystal structure obtained for the Oct1/Sox2/*FGF4* complex was compared to crystal structures of HMG-D (Murphy *et al.*, 1999), Lef1 (Love *et al.*, 1995) and Sry (Werner *et al.*, 1995) in complex with DNA (figure 4.3A). Sox2 generates a similar bend angle (90°) in DNA as reported for the other HMG/DNA complex structures (HMG-D: 111°; Lef1: 117°; Sry: 75°). α-Helix 1 and 2 form extensive contacts with the DNA in the widened minor groove and are involved in bending the DNA in all structures. However, the proteins show significant differences in how their C-terminal region interacts with the DNA molecule. In HMG-D, which binds DNA without sequence specificity, only α-helices 1 and 2 interact with the DNA. The C-terminus of the Lef1 HMG domain lies in the compressed major groove and stabilizes the bent DNA conformation. In the Sry/DNA structure, the C-terminal part is mainly disordered and is not positioned in the minor groove either.

4.3.3 The POU/HMG interface

In contrast to the other HMG structures, the C-terminus of the HMG domain (residues 68-79) in the Oct1/Sox2/*FGF4* complex is ordered and bound in an extended β-strand-like conformation in the compressed minor groove located between the HMG and the POU domain DNA binding sites (figure 4.3A). This C-terminal segment interacts with the DNA and also forms a protein-protein interface by

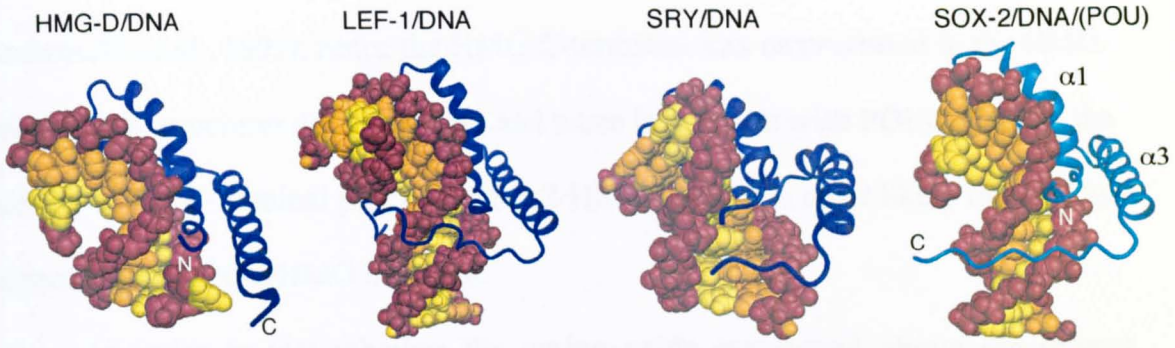
Figure 4.3

Comparison of different HMG domains

(**A**) Structures of different HMG domains from HMG-D (Murphy *et al.*, 1999), Lef1 (Love *et al.*, 1995), Sry (Werner *et al.*, 1995) and Sox2 (taken from the Oct1/Sox2/*FGF4* crystal structure) in complex with DNA. The DNA sugar-phosphate backbone is colored in red-brown and the bases belonging to different chains of the DNA molecule are depicted by different colors (yellow and orange). α-Helix 2 (α2) is behind the DNA molecule and, therefore, cannot be seen in this orientation.

(**B**) Multiple sequence alignment of HMG domains of several proteins from different organisms (y: yeast; c: *Caenorhabditis elegans*; d: *Drosophila melanogaster*; m: mouse; h: human). The order of the proteins on the list reflects their sequence similarity. HMG domains from the first six proteins (dHmgd-hHmg2b) are known to bind DNA in a non-sequence-specific manner, whereas the others (hLef1-hSox13) bind DNA sequence-specifically. Secondary structure elements of Sox2 from the Oct1/Sox2/*FGF4* crystal structure are shown beneath the alignment. Protein residues that are highly conserved are boxed in grey. It is noteworthy that the protein residues that play an important role in the ordering of the C-terminal region of the HMG from Sox2 (V3, R5, P6, H63, H67, P68, Y70, Y72, R75 and R76) are conserved in almost all members of the Sox family. However, these residues are divergent in HMG domains that bind DNA in a non-sequence specific manner. Residues P68, Y72 and P74, which play the most prominent role in positioning the HMG C-terminus into the compressed minor groove in the Oct1/Sox2/*FGF4* complex (cf. figure 4.4A), are colored in red.

A



B

```
          1.......10........20........30........40........50........60........70........80
dHmgd     DKPKRPLSAYMLWLNSARESIKRENPGIKVTEVAKRGGELWRAM--KDKSEWEAKAAKAKDDYDRAVKEFEANGGSSAAN
hSsrp1    NAPKRPMSAYMLWLNASREKIKSDHPGISITDLSKKAGEIWKGMSKEKKEEWDRKAEDARRDYEKAMKEYEGGRGESSKR
yNhp6a    NAPKRALSAYMFFANENRDIVRSENPDITFGQVGKKLGEKWKALTPEEKQPYEAKAQADKKRYESEKELYNATLA-----
yNhp6b    NAPKRGLSAYMFFANENRDIVRSENPDVTFGQVGRILGERWKALTAEEKQPYESKAQADKKRYESEKELYNATRA-----
hHmg1b    NAPKRPPSAFFLFCSEYRPKIKGEHPGLSIGDVAKKLGEMWNNTAADDKQPYEKKAAKLKEKYEKDIAAYRAKGKPDAAK
hHmg2b    NAPKRPPSAFFLFCSEHRPKIKSEHPGLSIGDTAKKLGEMWSEQSAKDKQPYEQKAAKLKEKYEKDIAAYRAKGKSEAGK
hLef1     PHIKKPLNAFMLYMKEMRANVVAECTLKESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYPGWSARDNYGKKK
yMtmc     ERTPRPPNAFILYRKEKHATLLKSNPSINNSQVSKLVGEMWRNESKEVRMRYFKMSEFYKAQHQKMYPGYKYQPRKNKVK
hSry      DRVKRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWKMLTEAEKWPFFQEAQKLQAMHREKYPNYKYRPRRKAKM
dDichaete GHIKRPMNAFMVWSRLQRRQIAKDNPKMHNSEISKRLGAEWKLLAESEKRPFIDEAKRLRALHMKEHPDYKYRPRRKPKN
hSox3     DRVKRPMNAFMVWSRGQRRKMALENPKMHNSEISKRLGADWKLLTDAEKRPFIDEAKRLRAVHMKEYPDYKYRPRRKTKT
hSox21    DHVKRPMNAFMVWSRAQRRKMAQENPKMHNSEISKRLGAEWKLLTESEKRPFIDEAKRLRAMHMKEHPDYKYRPRRKPKT
hSox14    DHIKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSEAEKRPYIDEAKRLRAQHMKEHPDYKYRPRRKPKN
mSox2     DRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKT
hSox1     DRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKVMSEAEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKT
mSox15    EKVKRPMNAFMVWSSVQRRQMAQQNPKMHNSEISKRLGAQWKLLGDEEKRPFVEEAKRLRARHLRDYPDYKYRPRRKSKN
hSox20    EKVKRPMNAFMVWSSAQRRQMAQQNPKMHNSEISKRLGAQWKLLDEDEKRPFVEEAKRLRARHLRDYPDYKYRPRRKAKS
hSox22    GHIKRPMNAFMVWSQHERRKIMDQWPDMHNAEISKRLGRRWQLLQDSEKIPFVREAERLRLKHMADYPDYKYRPRKKSKG
hSox4     GHIKRPMNAFMVWSQIERRKIMEQSPDMHNAEISKRLGKRWKLLKDSDKIPFIREAERLRLKHMADYPDYKYRPRKKVKS
hSox11    GHIKRPMNAFMVWSKIERRKIMEQSPDMHNAEISKRLGKRWKMLKDSEKIPFIREAERLRLKHMADYPDYKYRPRKKPKM
mSox17    SRIRRPMNAFMVWAKDERKRLAQQNPDLHNAELSKMLGKSWKALTLAEKRPFVEEAERLRVQHMQDHPNYKYRPRRRKQV
hSox18    SRIRRPMNAFMVWAKDERKRLAQQNPDLHNAVLSKMLGKAWKELNAAEKRPFVEEAERLRVQHLRDHPNYKYRPRRKKQA
hSox10    PHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNESDKRPFIEEAERLRMQHKKDHPDYKYQPRRRKNG
hSox9     PHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNESEKRPFVEEAERLRVQHKKDHPDYKYQPRRRKSV
hSox8     PHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLSESEKRPFVEEAERLRVQHKKDHPDYKYQPRRRKSA
cCog2     NHIKRPMNAFMVWARDERRKILKAYPDMHNSNISKILGSRWKGMSNSEKQPYYEEQSRLSKLHMEQHPDYRYRPRPKRTC
mSox6     PHIKRPMNAFMVWAKDERRKILQAFPDMHNSNISKILGSRWKSMSNQEKQPYYEEQARLSKIHLEKYPNYKYKPRPKRTC
hSox5     PHIKRPMNAFMVWAKDERRKILQAFPDMHNSNISKILGSRWKAMTNLEKQPYYEEQARLSKQHLEKYPDYKYKPRPKRTC
hSox13    SHIKRPMNAFMVWAKDERRKILQAFPDMHNSSISKILGSRWKSMTNQEKQPYYEEQARLSRQHLEKYPDYKYKPRPKRTC
                    α1                  α2                  α3
```

contacting a loop of the Oct1 POU_S domain between α-helices 1 and 2. The crystal structure reveals only one sequence-specific interaction between the two domains, a salt bridge between R75 (HMG) and D29 (POU_S), which could provide a rationale for the observed indiscriminate nature of POU/HMG/DNA complex formation (Ambrosetti *et al.*, 1997). Since the HMG C-terminus was unstructured in the HMG-protein/DNA structures solved to date and since it interacts with POU_S of Oct1, the ordering of the C-terminal part of the Sox2 HMG domain is most likely induced by the presence of a POU/HMG interface.

In order to test whether the amino acids mentioned above are indeed necessary for POU-Sox interaction on *FGF4*, they were mutated by site directed mutagenesis through which their charges were reversed. In the HMG domain, the positively charged arginine (R75) was mutated into a negatively charged, bulky glutamic acid (E). In the POU domain the negatively charged aspartic acid (D29) was mutated into a positively charged arginine and isoleucine (I21), which forms a van-der-Waals interaction with the HMG domain of Sox2, was mutated into a bulky tyrosine (Y). EMSAs show that the unmutated POU1 and HMG domain of Sox2 (wt) bind to *FGF4* together (figure 4.2C lane 2). As soon as either the I21Y,D29R Sox2 mutant (m1) is combined with the unmutated POU1, or the R75E POU1 mutant (mut) with the unmutated Sox2, ternary complex formation is drastically reduced (lanes 3 and 5). The two mutated proteins do not dimerize on *FGF4* at all (lane 6).

The EMSA in figure 4.2C displays the importance of R75, which is at the very C-terminus of the HMG domain. This supports the hypothesis, based on the comparison of the ternary structure with other HMG-protein/DNA ternary complexes, that the ordering of the C-terminal part of the Sox2 HMG domain is induced by the presence of a POU/HMG interface.

A multiple sequence alignment (figure 4.3B) reveals that this HMG C-terminus (residues 68-79) is virtually identical among all Sox members of the HMG family, but unrelated in sequence to members of other HMG subfamilies. This suggests that heterodimer interface formation via the HMG domain C-terminus is a property limited to the Sox subgroup of HMG proteins. Moreover, the complementary POU_S surface patch is also highly conserved among POU factors (Herr and Cleary, 1995). The C-terminus of the Sox2-HMG domain, which is presumably unstructured in the absence of an interacting protein partner, is likely to be a major contributor to ternary complex formation, since the interaction of this portion of the protein with the compressed minor groove increases the HMG protein/DNA surface by about one third of the total (420 $Å^2$/1350 $Å^2$). This implies that upon HMG/POU interaction and thus ordering of the HMG C-terminus, the HMG/DNA interaction surface becomes bigger and thus stronger, just by unspecific van-der-Waals bonds. The formation of specific hydrogen bonds or salt bridges would further enhance the strength of the interaction – on top of the increase caused by proximity in space. This is the case for tyrosine (Y72) of the HMG C-terminus, which interacts sequence-specifically with the DNA via a hydrogen bond in the Oct1/Sox2/*FGF4* ternary complex (figure 4.2B). Sox2 is able to bind to DNA on its own, but with a significantly lower affinity compared to binding to DNA as part of a ternary complex with POU or Pax proteins. This observation is in agreement with the supposition that Sox proteins are converted into high-affinity ligand binders in the presence of other DNA binding protein partners (Kamachi *et al.*, 2000). The ordering of the C-terminus and hence an increase in protein/DNA interaction surface could be one explanation for cooperativity of POU/HMG interaction due to a strong POU/HMG protein-protein interface.

4.3.4 Homology modeling of Oct4/Sox2/DNA ternary complexes

The Oct1/Sox2/*FGF4* crystal structure allowed my collaborator Attila Reményi (EMBL Hamburg) to generate a reliable Oct4/Sox2/*FGF4* model based on homology (Vriend, 1990), and to build an Oct4/Sox2/*UTF1* ternary complex model (figure 4.4A). The latter was created by using the experimental structure of the Oct1/Sox2/*FGF4* complex and keeping the Sox2-HMG domain and the Oct1/Oct4-POU domains (POU1 and POU4) at their cognate DNA sequences. Because Sox2 and POU factors bind DNA in a highly sequence-specific manner, their positioning on composite DNA sites can be inferred reliably. The new model is based on a rotational movement of the HMG domain towards POU_S, and suggests a second POU-HMG interface, which is different from that on *FGF4*. In both arrangements, the same surface patch of the POU_S subdomain is used. However, according to the *UTF1*-model, the Sox2-HMG domain interacts via a segment of α-helix 3 (K57-M64) on *UTF1* instead of its C-terminus as is the case on *FGF4*. The HMG interaction surface thus changes from an extended β-strand-like structure to an α-helical interface.

4.3.5 Sox2 has two surface patches for interaction with Oct4

To test whether these two distinct protein interfaces indeed exist on the Sox2-HMG protein, I introduced two supposed *UTF1*-specific mutations in its C-terminal region: K57E, R60E (m2) and R60E, M64E (m3). None of the mutations had a significant impact on Sox2/DNA interaction, as the individual proteins bound to DNA similar to the respective wild-type protein.

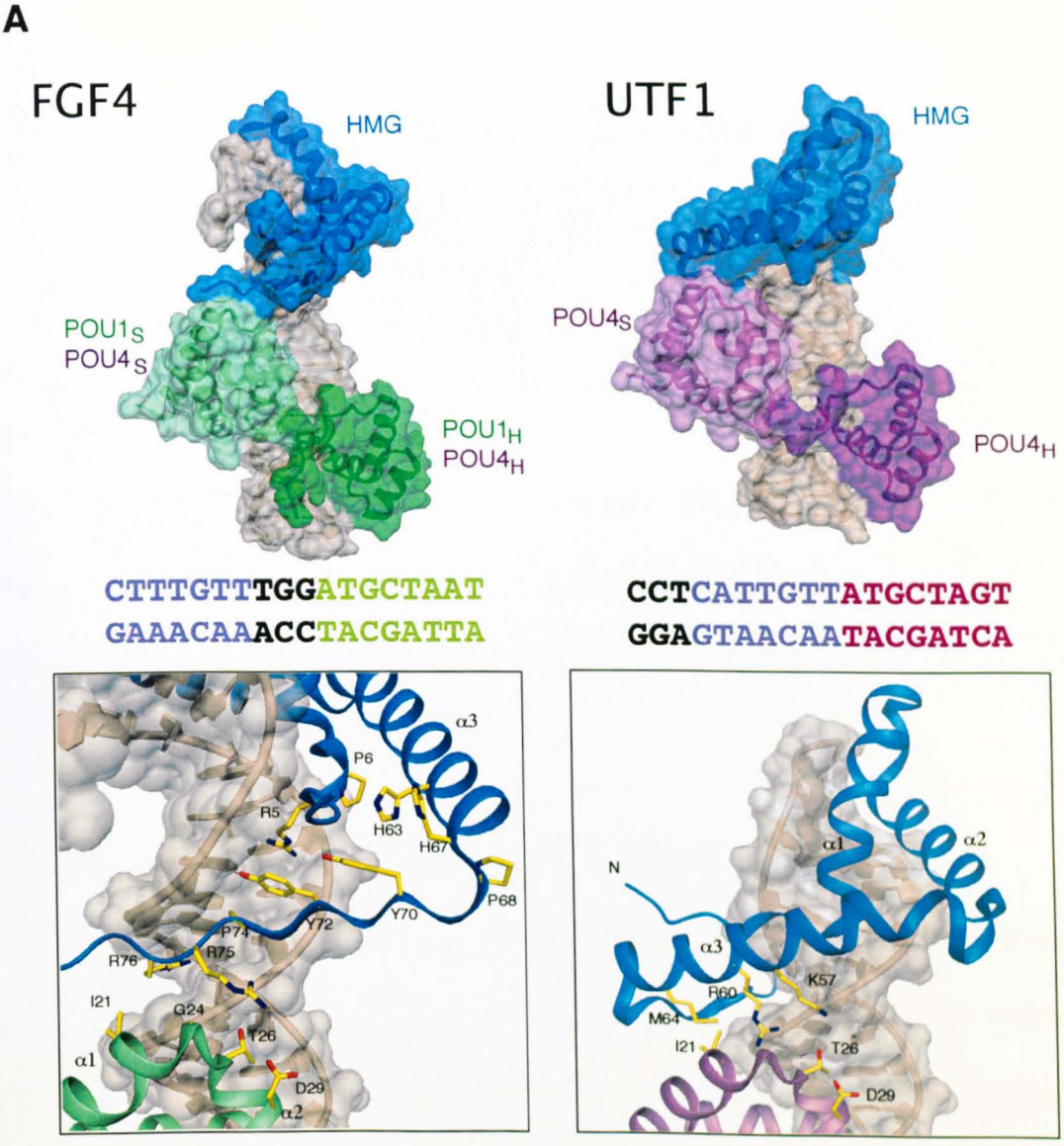
The proteins were compared to the unmutated HMG and the *FGF4*-specific m1 mutant (R75E) in terms of heterodimer formation with Oct4 in EMSAs (figure 4.4B). The three mutants showed differential ternary complex formation with the

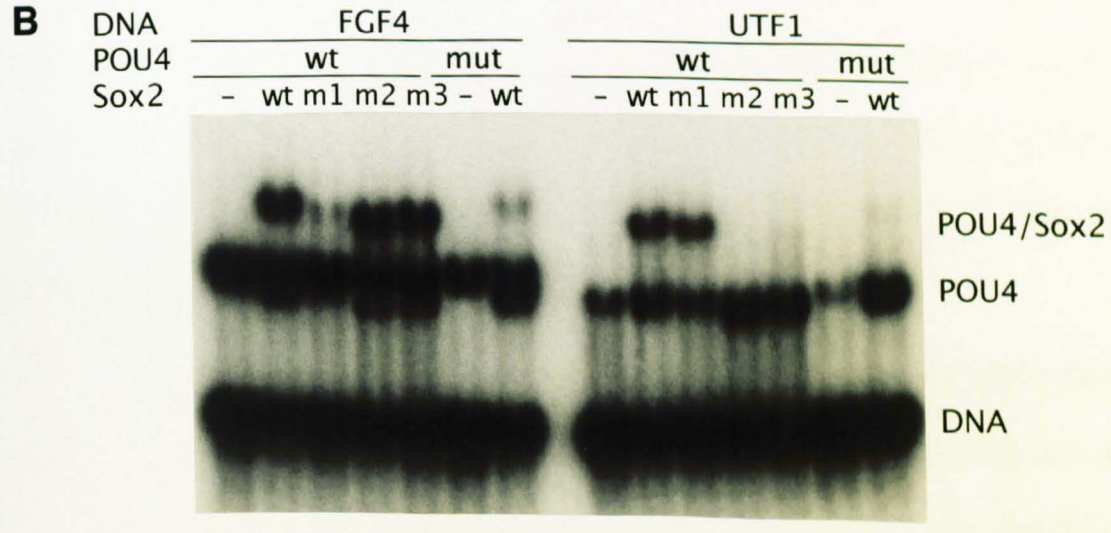
Figure 4.4

Comparison of POU/HMG complexes formed on *FGF4* and *UTF1*

(A) Upper panel: model of POU/HMG/*FGF4* (POU domain of Oct1 or Oct4) (left), compared to the model of Oct4-POU/HMG/*UTF1* (right). The figure illustrates that different spacing between the binding sites for the POU and HMG domains within *FGF4* and *UTF1* elements causes formation of different heterodimeric interfaces. Lower panel: close-up views of the HMG/POU_S interfaces on *FGF4* and *UTF1*. The POU/HMG/*UTF1* model suggests involvement of α-helix 3 instead of the C-terminus of the HMG domain to form the HMG-POU_S interface, while the same surface patch of the POU_S domain appears to be involved in both interfaces. The DNA molecules are depicted with a transparent surface and are colored in grey.

(B) To validate the homology models in the upper panel, mutations in Sox2-HMG were designed to selectively interfere with ternary complex formation on the *FGF4* (m1) or on the *UTF1* element (m2 and m3); m1: R75E; m2: K57E,R60E; m3: R60E,M64E. mut: mutant version of Oct4 POU (I21Y,D29E). In agreement with both models, m1 specifically impaired heterodimerization on *FGF4*, while m2 and m3 specifically abrogated heterodimerization on *UTF1*, whereas POU4 mut affected both.


A
FGF4
HMG
POU1S
POU4S
POU1H
POU4H
UTF1
HMG
POU4S
POU4H
CTTTGTTTGGATGCTAAT
GAAACAAACCTACGATTA
CCTCATTGTTATGCTAGT
GGAGTAACAATACGATCA
α3
P6
R5
H63
H67
P68
Y70
Y72
P74
R75
R76
I21
Q24
T26
α1
D29
α2
N
α1
α2
α3
K57
R60
M64
I21
T26
D29


B
DNA
POU4
Sox2
FGF4
wt
mut
- wt m1 m2 m3 - wt
UTF1
wt
mut
- wt m1 m2 m3 - wt
POU4/Sox2
POU4
DNA

FGF4 and *UTF1* elements. Although both *FGF4* and *UTF1* mediated assembly of a ternary complex with the POU domain of Oct4 and wild-type Sox2-HMG, heterodimerization with either m2 or m3 on *UTF1* and with m1 on *FGF4* was selectively compromised. As established earlier, I21 and D29 of the POU domain play a prominent role in protein-protein interaction with HMG in the POU/HMG/*FGF4* complex (figure 4.2C). According to the POU/HMG/*UTF1* model the same POU surface patch implicated in heterodimerization on *FGF4* is also involved on *UTF1*. In agreement with this observation, the I21Y, D29R double mutation in POU4 disrupted the ternary complex formation on both elements. In summary, these results validate the proposed models of Oct4/Sox2/DNA complexes, in which distinct surface patches of the HMG domain of Sox2 interact with the same surface patch of the POU domain of Oct factors on *FGF4* and *UTF1*.

4.3.6 Sox2 interacts with Oct4 and Pax6 via the same interface

Sox proteins can also establish direct functional partnerships with members of the Pax transcription factor family (Kamachi *et al.*, 2000). Pax factors contain a conserved 128-amino-acid DNA binding 'paired domain' and a C-terminal transactivation domain. They play critical roles in mammalian development and oncogenesis (Mansouri *et al.*, 1996). Sox2 specifically interacts with Pax6 on the *DC5* enhancer (Kamachi *et al.*, 2001). This enhancer controls the expression of the δ-crystallin gene, which plays a pivotal role in eye development during late embryogenesis (Kamachi *et al.*, 2001).

To test whether Pax6 interacts with one of the two interfaces on Sox2 introduced earlier, an EMSA of the *DC5* element and Pax6 was performed with mutants m1, m2 and m3 in comparison to the wild-type HMG (figure 4.5). Figure 4.5A depicts the

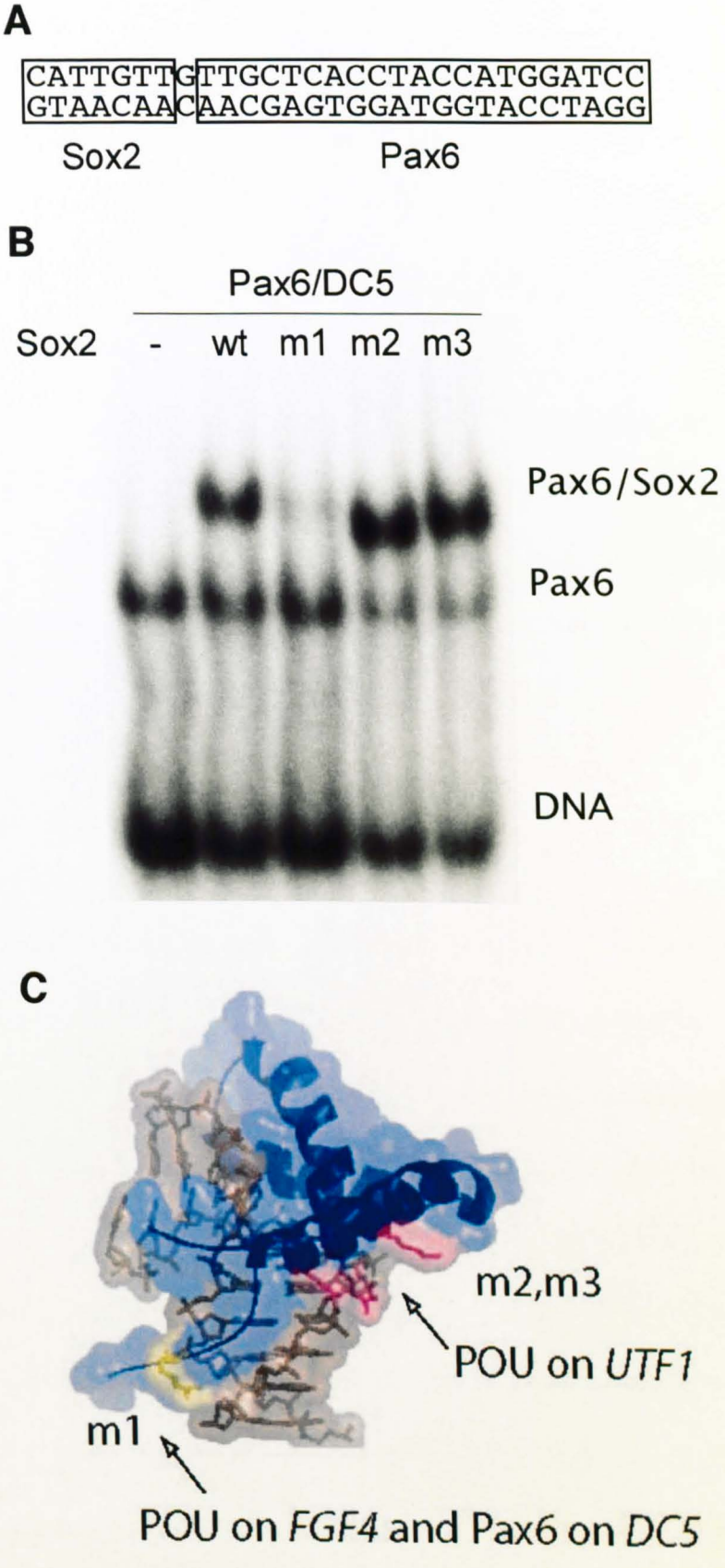
Figure 4.5

Sox2-Pax6 on *DC5* and Sox2-Oct4 on *FGF4* use the same HMG interface

(A) Schematic representation of the Sox2 and Pax6 DNA binding site in the *DC5* enhancer.

(B) EMSA showing that the same mutation in HMG, which interferes with POU/Sox2 on *FGF4* (m1), also impairs complex formation with Pax6 on *DC5*, indicating that Sox2 uses the same interface for ternary complex assembly on these elements.

(C) Schematic presentation showing the two non-overlapping interfaces of Sox2 required for protein-protein interactions. The position of R75 (m1) and the positions of K57, R60 and M64 (m2 and m3) are shown on a transparent surface of the Sox2/DNA complex from the Oct1/Sox2/*FGF4* crystal structure in yellow and magenta, respectively.


A
CATTGTTGTTGCTCACCTACCATGGATCC
GTAACAACAACGAGTGGATGGTACCTAGG
Sox2
Pax6
B
Pax6/DC5
Sox2 - wt m1 m2 m3
Pax6/Sox2
Pax6
DNA
C
m2,m3
POU on *UTF1*
m1
POU on *FGF4* and Pax6 on *DC5*

sequence of the *DC5* element and the Sox2 and Pax6 DNA-binding sites therein. As can be seen in the EMSA (figure 4.5B), in contrast to the HMG mutant m1, wild type HMG as well as mutants m2 and m3 heterodimerize with Pax6 on *DC5*. This experiment revealed that the same HMG domain mutation, R75E (m1) that specifically interferes with Oct/Sox2/*FGF4* binding also abrogates Pax6/Sox2/*DC5* ternary complex formation, whereas the *UTF1*-specific mutants (m2, m3) interacted with Pax6. This finding suggests that the same Sox2 interface (the C-terminal region of HMG) is required for heterodimer formation with Oct4 on *FGF4* and with Pax6 on *DC5*, although these two Sox2 partners are members of different transcription factor families and, as such, unrelated in sequence and structure. A schematic representation of the non-overlapping areas of the HMG domain involved in the interaction with the POU domain of Oct factors and Pax6 is presented in figure 4.5C.

4.4 Discussion

4.4.1 POU and Sox proteins establish combinatorial codes during development

Sox genes are expressed in various phases of embryonic development and cell differentiation. They are recognized as key players in the determination of cell fate (Pevny and Lovell-Badge 1997). Because HMG domains of Sox proteins are similar to each other in their DNA sequence preference (Mertin *et al.*, 1999) and in their DNA-bending activity (Kamachi *et al.*, 1999), it remains elusive how they are capable to regulate different target genes. Assembly with unrelated transcriptional regulator proteins, however, provides a plausible explanation of how they can distinguish their targets as well as act in a cell-specific fashion (Kamachi *et al.*, 2000). Partnering with members of the POU and Pax family of transcription factors, as shown in this chapter, may provide paradigm examples.

Interactions between various Sox and POU factors provide the best characterized examples for Sox partnerships with members of other transcription factor families. There is a substantial number of well-characterized examples for this alliance in mouse (Oct4/Sox2, Oct6/Sox10, Brn1,2/Sox11) and in the fruit fly (Drifter/Dichaete) underlying their versatile involvement in different biological functions (Yuan *et al.*, 1995; Botquin *et al.*, 1998; Nishimoto *et al.*, 1999; Kuhlbrodt *et al.*, 1998a,b; Soriano and Russell 1998). The Oct4/Sox2 partnership, for example, plays a fundamental role in determining the pluripotent cell state in early embryos (Yuan *et al.*, 1995; Nishimoto *et al.*, 1999). In contrast, Oct6/Sox10 and Brn1,2/Sox11 pairs have been shown to be involved in glial cell development in mouse (Kuhlbrodt *et al.*, 1998b). The Drifter/Dichaete cooperation induces development of the central nervous system in the fruit fly (Soriano and Russell 1998). This latter finding suggests that POU/HMG partnerships during development might be evolutionarily conserved from invertebrates to mammals.

The list of Sox-interacting partners, however, is not limited to the POU and Pax families of transcription factors. Sox9, for example, activates two of its target genes in conjunction with yet other transcription factors (Kamachi *et al.*, 2000). It regulates *Col2a1*, encoding type II collagen, during chondrogenesis in cartilage tissue, and the anti-Müllerian hormone gene (*AMH*) during male sex determination in the genital ridge. In both cases the proper expression pattern of the target gene requires the Sox binding site in proximity to another conserved DNA binding motif on the regulator sequence. Activation of expression requires the concomitant binding of Sox9 with an additional transcription factor. In the case of *Col2a1* the factor has not been identified yet, for *AMH* it has been established as SF1, a member of the orphan nuclear receptor family (De Santa Barbara *et al.*, 1998).

The work in this chapter sheds light onto the molecular mechanism of POU and Sox partnering on certain enhancers. It will be interesting to see if the principles unraveled here can be directly applied to other POU and Sox factor pairs or even – similarly to the resemblance of POU/Sox and Pax/Sox interaction – for partnerships with further protein families.

4.4.2 *In vivo* importance of differential POU/HMG interaction on *FGF4* and *UTF1*

As *FGF4* and *UTF1* are differentially expressed during early mouse development, insight might be gained by comparing the activity of these two genes to the levels of their regulators: Sox2 and Oct4. The cell lines ES, F9 EC and P19 EC are the *in vitro* counterpart of early stem cell types of different embryonal stages and, as such, provide useful cell culture models (Yeom *et al.*, 1996). Oct4 protein levels are similar in the three cell lines, while Sox2 protein levels vary (figure 4.6) (Yeom *et al.*, 1996; Botquin *et al.*, 1998). An interesting finding is that both, Sox2 and FGF4, have the highest expression level in ES cells followed by F9 EC and then P19 EC cells, whereas *UTF1* levels are similar to each other in all three cell lines (figure 4.6).

One formidable hypothesis is that the different *FGF4* and *UTF1* activities during development are related to differences in the cooperativity of POU and HMG domain interactions on their respective elements. Differential cooperativity may provide a rationale for how the *FGF4* and *UTF1* genes respond to varying amounts of Oct4 and Sox2 proteins present during early development. Due to a higher level of cooperativity, Oct4 may require comparably little Sox2 to heterodimerize and activate *UTF1* expression. Conversely on *FGF4*, heterodimers and thus increased gene expression levels may depend on higher amounts of Sox2. The comparative titration experiments (figure 4.1A) demonstrate that a weaker concentration of the HMG

Figure 4.6

Expression levels of gene products in three different cell lines representing different cell types during early embryonic development

= indicates that the level of a gene product is similar between the cell lines.

The comparison is based on the following references:

Oct4: [1]Botquin *et al.*, 1998

Sox2: [2]Yuan *et al.*, 1995; [3]Dailey *et al.*, 1994

FGF4: [4]Schoorlemmer and Kruijer 1991

UTF1: [5]Okuda *et al.*, 1998

	ES cells	F9 EC cells	P19 EC cells
Oct4	= [1]	= [1]	= [1]
Sox2	high [2]	medium [2,3]	low [3]
FGF4	high [4]	medium [4]	low [4]
UTF1	= [5]	= [5]	= [5]

domain of Sox2 permits heterodimerization with Oct4 on *UTF1* than on *FGF4*.

The different degrees of Sox2/Oct4 cooperativity on the regulatory elements *in vitro* are in congruence with the sequential up- and down-regulation of *UTF1* and *FGF4* during development, but these associations need to be tested *in vivo*. A stringent functional test would require specific point mutations to be introduced into the *Oct4* and *Sox2* genes. This approach has become practical only after the construction of reliable three-dimensional models presented in this study, and their verification by testing the involvement of specific amino acid residues of the HMG domain of Sox2 and the POU domain of Oct4 in heterodimerization by mutating them and assessing their capability to interact with each other on the DNA elements concerned.

4.4.3 Comparison of POU/POU homo- and POU/HMG heterodimerization

In collaboration with two former members of the laboratory in which I performed the research for this dissertation, we have reported that Oct factors are capable of homodimerization on two functionally distinct elements, termed PORE and MORE (Tomilin *et al.*, 2000; Reményi *et al.*, 2001). This binding is mediated by separate dimerization surface patches of the conserved POU domain. The POU domains of all factors tested can dimerize via the two different surface patches on the PORE and the MORE without steric clashes. Even though several interacting amino acids are not conserved from one protein to the other, their characteristics are similar enough to perform the same function (Reményi *et al.*, 2001). On the MORE, for example, the C-terminal residues of POU_H of one molecule dock onto the loop region between α-helices 3 and 4 of POU_S of the other molecule. The side chain of isoleucine 159 of Oct1 fits into a hydrophobic cavity of the POU_S subdomain, which

forms a "knob-in-the-hole" structure. In Oct4 its counterpart serine 159 could play a pivotal role in the MORE interface: it could be involved in hydrogen bond formation with glutamine 6 from POU_S of the other Oct4 molecule. The modeling, along with previous biochemical data (Botquin *et al.*, 1998; Tomilin *et al.*, 2000), suggests that the observed interface swapping is a property widely shared in the POU family. Clearly, the transcriptional properties of each POU factor and its regulatory mechanisms need to be better characterized, in order to discover to what extent the potential of interface-domain swapping is generally used to acquire differential transcriptional activity.

Interestingly, Sox2 also contains two functionally and structurally distinct protein interaction surfaces. As is the case for POU dimers, the distance between the domain binding sites within the DNA motif is critical for selecting between different interfaces of Sox2. This property is very likely to be instrumental in creating various multi-protein/DNA complexes with distinct biochemical properties. The differences *in vitro*, such as varying amount of cooperativity in complex formation, could result in distinct functional properties *in vivo*, such as varying thresholds of transcription factors required for activation or different extents of transcript level production. Furthermore, the different quaternary arrangement of transcription factor/DNA complexes could serve as the basis for differential recruitment of specific coregulators, as has been shown for POU factor homodimerization of Oct1 and Pit1 on different elements (Scully *et al.*, 2000; Tomilin *et al.*, 2000; Reményi *et al.*, 2001; Reményi *et al.*, 2002).

4.4.4 DNA mediated interaction surface swapping as a general model

The data presented in this chapter supports the emergence of a novel integrative approach to define the principles underlying differential complex

formation on DNA. Specifically, it subsumes that various combinations of transcription factors and their coregulators are possible, along with the potential of some of these proteins to interchange their quaternary DNA-mediated arrangements via one of multiple surfaces capable of protein-protein interactions (figure 4.7). Therefore, a certain dimerization surface patch appears to be adept at mediating Velcro-like surface interactions with different interacting protein partners in a versatile fashion. As such, this study provides insight into the adaptive mechanisms used by transcription factors to assume a regulatory stronghold on various complex processes during mammalian development.

Figure 4.7

DNA sequence dependent protein interaction surfaces of Sox2 and Oct factors

Interface areas involved in protein-protein interaction in a DNA-sequence specific manner are marked on the sequence of Sox2, Oct1 and Oct4 (non-variant residues of Oct4 and Oct1 shown with a dot). Several of these areas interact with several other protein partners resulting in various combinations of homo- and heterodimers. The loop region between the first and second α-helices in the POU domain, for example, is involved in interactions with Sox2 on the *FGF4* and *UTF1* elements as well as in the homodimerization with a second Oct factor on the PORE element. Also, the C-terminal region of the POU domain interacts not only with a second Oct factor on the MORE but – in the case of Oct1 – also interacts with a coactivator (OBF1) on the PORE as well as on the octamer motif (Chasman *et al.*, 1999; Reményi *et al.*, 2001; Chapter 3).

Pax6/DC5
Oct4/UTF1 Oct4/FGF4
Sox2 (HMG)
1 80
DRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKT
α1 α2 α3
Oct1 and Oct4 (POU)
1 75
101 160
EEPSDLEELEQFAKTFKQRRIKLGFTQGDVGLAMGKLYGNDFSQTTISRFEALNLSFKNMCKLKPLLEKWLNDAE
DMKALQK.....LL..K..T..Y..A....TL.V.F.KV.....C.....Q..L.....R.....VEE.D
Oct1
Oct4
RRKKRTSIETNIRVALEKSFLENQKPTSEEITMIADQLNMEKEVIRVWFCNRRQKEKRIN
.KR.....NRV.WS..TM..KCP..SLQQ..H..N..GL..D.V.....G..SS
α1 α2 α3 α4
..... linker .....
α1 α2 α3
Oct/PORE
Sox2/FGF4
Sox2/UTF1
Oct/MORE
OBF1/PORE
OBF1/octamer motif
Oct/PORE
Oct/MORE
OBF1/PORE
OBF1/octamer motif

Chapter 5

Sox2-Oct4 dimer modulates PORE activity via remote Sox2 site

5.1 Abstract

The transcription factors Oct4 and Sox2 regulate transcription of *Osteopontin, FGF4* and *UTF1*. The first intron of *Osteopontin* contains a Sox-binding site and a unique PORE with two overlapping elements, binding Oct4 either as a monomer or a dimer. Whereas both factors synergistically activate *FGF4* and *UTF1*, Sox2 interferes with Oct4-mediated activation of *Osteopontin*. This chapter reveals that Sox2-specific repression depends on the upstream Sox site and an intact PORE, although neither the Sox nor the PORE sites are negative elements on their own. The Oct4 dimer mediates high levels of activation via the PORE, which is reduced if Sox2 can communicate with the Oct4 monomer. The results presented in this chapter indicate that the 34 bp distant Sox site serves as a recruiting site helping Sox2 proteins enter and bridge the gap to the Oct4 monomer. This results in a replacement of the Oct4-Oct4 homodimer by an inactive Sox2-Oct4 heterodimer and consequently a net reduction of *Osteopontin* enhancer activity.

5.2 Introduction

Development is defined by a series of genetic and epigenetic events that generate the adult body from the unicellular zygote. In contrast to the complex regulatory network that is established during embryogenesis, only a small number of transcription factors interact with each other and cofactors to enable this differential gene expression during development.

Members of the POU and Sox families play critical roles in diverse developmental processes, providing a model for investigating the mechanisms of gene activation by transcription factor complexes (Ryan and Rosenfeld, 1997; Wegner, 1999). For example, most recently, the synergistic activation of the Sox17 promoter

by the Oct4-like POU factor SPG and the Sox-related Casanova was shown to be required for endoderm formation in zebrafish (Reim *et al.*, 2004). To date, more than 20 metazoan POU factors and over 30 Sox members have been identified (Ryan and Rosenfeld, 1997; Bowles *et al.*, 2000). The members of these two families are involved in diverse developmental processes and, with the exception of the ubiquitous Oct1, are differentially expressed during embryogenesis.

Sox factors (reviewed in Wegner, 1999), a subgroup of the HMG-box superfamily, all recognize a similar DNA binding motif, C(A/T)TTG(A/T)(A/T), and induce a sharp bend in the double helix. Sox proteins may function both as conventional transcription factors and as architectural proteins organizing local chromatin structure and assembling other DNA-bound transcription factors into biologically active multiprotein complexes (Ferrari *et al.*, 1992; Giese *et al.*, 1992).

Oct4 and Sox2 are representative transcription factors of these two families, and can be used to study the combinatorial crosstalk between two developmentally relevant factors. Oct4 is expressed in the early mouse embryo and is later restricted to the germline. The gene is active in a number of distinct embryonic cell lines, including embryonic stem (ES) and embryonic carcinoma (EC) cell lines (for reviews see Ovitt and Schöler, 1998; Pesce *et al.*, 1999). Oct4 is downregulated upon the differentiation of the embryonic cell lines (Schöler *et al.*, 1989a) and it is required for the maintenance of embryonic cell potency (Minucci *et al.*, 1996; Yeom *et al.*, 1996). Furthermore, it may serve as a determinant of the germ cell fate by preventing their differentiation to a somatic cell phenotype during gastrulation. Sox2 is coexpressed with Oct4 in early mouse embryogenesis, ES and EC cells (Avilion *et al.*, 2003).

Oct4 and Sox2 are considered to be involved in the transcriptional regulation of various genes that are differentially expressed during early mouse development,

including *FGF4*, *Osteopontin* (*OPN*), *UTF1*, *Sox2* and *Fbx15* (Yuan *et al.*, 1995; Botquin *et al.*, 1998; Nishimoto *et al.*, 1999; Tomioka *et al.*, 2002; Tokuzawa *et al., 2003*). The regulatory regions of these three genes contain binding sites for Oct- and Sox-factors located within the 5' promoter of *UTF1*, the 3' UTR of *FGF4*, the first intron of *OPN*, and the 3' enhancer of *Sox2*.

It appears that the arrangement of their binding sites dictates whether Oct4 and Sox2 interact synergistically or antagonistically. In *FGF4* and *UTF1* the Oct4 and Sox2 binding sites are 3 base pairs and 0 base pairs apart, respectively, with both factors cooperating in a synergistic manner (Nishimoto *et al.*, 1999; Yuan *et al.*, 1995). In *FGF4*, a specific spatial arrangement between both binding sites enables Oct4 and Sox2 to form a ternary complex with precise stereospecific requirements (Ambrosetti *et al.*, 1997, Ambrosetti *et al.*, 2000, Chapter 4). Moreover, Oct4 and Sox2 can bind to each other, *in vitro*, in the absence of DNA (Ambrosetti *et al.*, 1997).

OPN is expressed in various tissues and cell types, including ES and EC cells (Denhardt *et al.*, 1995; Botquin *et al.*, 1998). It is a candidate target gene of Oct4, and contains an EC cell-specific enhancer element that is regulated by Oct4 and Sox2 (Botquin *et al.*, 1998). The spatial arrangement of the octamer and Sox recognition sites in the first intron of the *OPN* gene is different from their configuration in the *FGF4* and *UTF1* enhancers. The PORE site, on which Oct4 can bind as a monomer or homodimer, was found to be both necessary and sufficient for strong transcriptional activation of the *OPN* enhancer. In contrast, Sox2 repressed the Oct4 stimulatory activity via a Sox element located 34 base pairs upstream of the PORE (Botquin *et al.*, 1998). These results are in dissonance with those of others regarding the synergistic

function of Sox2 and Oct4 in the transcriptional regulation of genes (Yuan *et al.*, 1995; Nishimoto *et al.*, 1999).

Based on these contrasting findings, I sought to understand how Sox2 interferes with Oct4-mediated transcriptional activation. I performed the experiments for this study in close collaboration with Valérie Botquin, a former member of the laboratory I conducted the research for this thesis in. The intact PORE as well as the upstream Sox site is required for Sox2 repression. The data presented in this chapter suggest that repression occurs by Sox2 binding to the Sox site and thereby facilitating Sox2 heterodimer formation with Oct4 on the downstream PORE. Sox2-mediated repression occurs as a consequence of the occupancy of the ternary complex on the PORE, thereby preventing Oct4 homodimer formation.

5.3 Results

5.3.1 Cross-talk between Oct4 and Sox2 in embryonic cell lines

Previously, Botquin *et al.* (1998) identified the PORE in *OPN* as a POU factor dimerization site, containing an octamer motif. A point mutational analysis showed that the Oct4 dimer elicits higher transcriptional activity than the monomer. Oct4 and Sox2 are thought to regulate *UTF1, FGF4* and *OPN* (Yuan *et al.*, 1995; Botquin *et al.*, 1998; Nishimoto *et al.*, 1999). Their binding sites are adjacent in *UTF1*, 3 base pairs apart in *FGF4,* and 41 base pairs in *OPN*. The Oct4/Sox2 heterodimer activates transcription of *FGF4* and *UTF1*, but Sox2 represses *OPN* enhancer activity. This chapter is aimed at determining how Sox2 interferes with Oct4-mediated dimer activity.

In order to assess the modality of Sox2 interference with Oct4 activation of the *OPN* enhancer, hexamers of different binding elements were inserted upstream of a

minimal herpes simplex virus *thymidine kinase* promoter (HSV *tk*) driving the expression of the *luciferase* (*luc*) reporter gene. The **SDM** oligonucleotide contains the wild-type sequence found in *OPN*. S, D, and M denote the three recognition elements in the *OPN* intron from 5' to 3' (**S**: canonical Sox2 recognition element; **D**: POU Dimer binding site; **M**: the octamer motif, a POU Monomer binding site within the dimer site; referred to as OS in Botquin *et al.*, 1998). Lower case letters (s, d and m) indicate mutated elements that specifically impair Sox2, Oct4 dimer, and Oct4 monomer binding *in vitro*, respectively (Botquin *et al.*, 1998).

The activities of the reporter plasmids were determined in two embryonic cell lines (F9 EC and R1 ES) expressing both Oct4 and Sox2 (figure 5.1A, B). The activity of the wild-type **SDM** reporter, although high, was 4.3- and 7.5-fold lower than that of the mutated **sDM** reporter in F9 EC and R1 ES cell lines, respectively. A lower activity in R1 ES cells correlate with a higher level of Sox2 protein in these cells compared to F9 EC (Chapter 4; Botquin *et al.*, 1998).

Poor induction of reporter activity was observed when the POU dimer site had been mutated and only the monomer site was left intact (compare **SdM** and **sdM** to *tk* in figures 5.1A and B). In both cases octamer activity was moderately higher when the Sox element was left intact (not obvious in the figure due to scale). Whereas the dimer mutations were almost inactive, the monomer mutants **SDm** and **sDm** were highly active, indicating that transcriptional activation is due to the dimer. Like for the dimer mutants, the monomer mutant with the intact Sox site (**SDm**) was slightly more active than the one with a mutant Sox site (**sDm**). These results showed that the Sox element, *per se*, is not a negative element, as it neither interfered with Oct4 monomer nor with dimer activity. If Sox2 acted as a general repressor of reporter gene activity, the **sDm** and **sdM** activation levels would have been higher than those

Figure 5.1

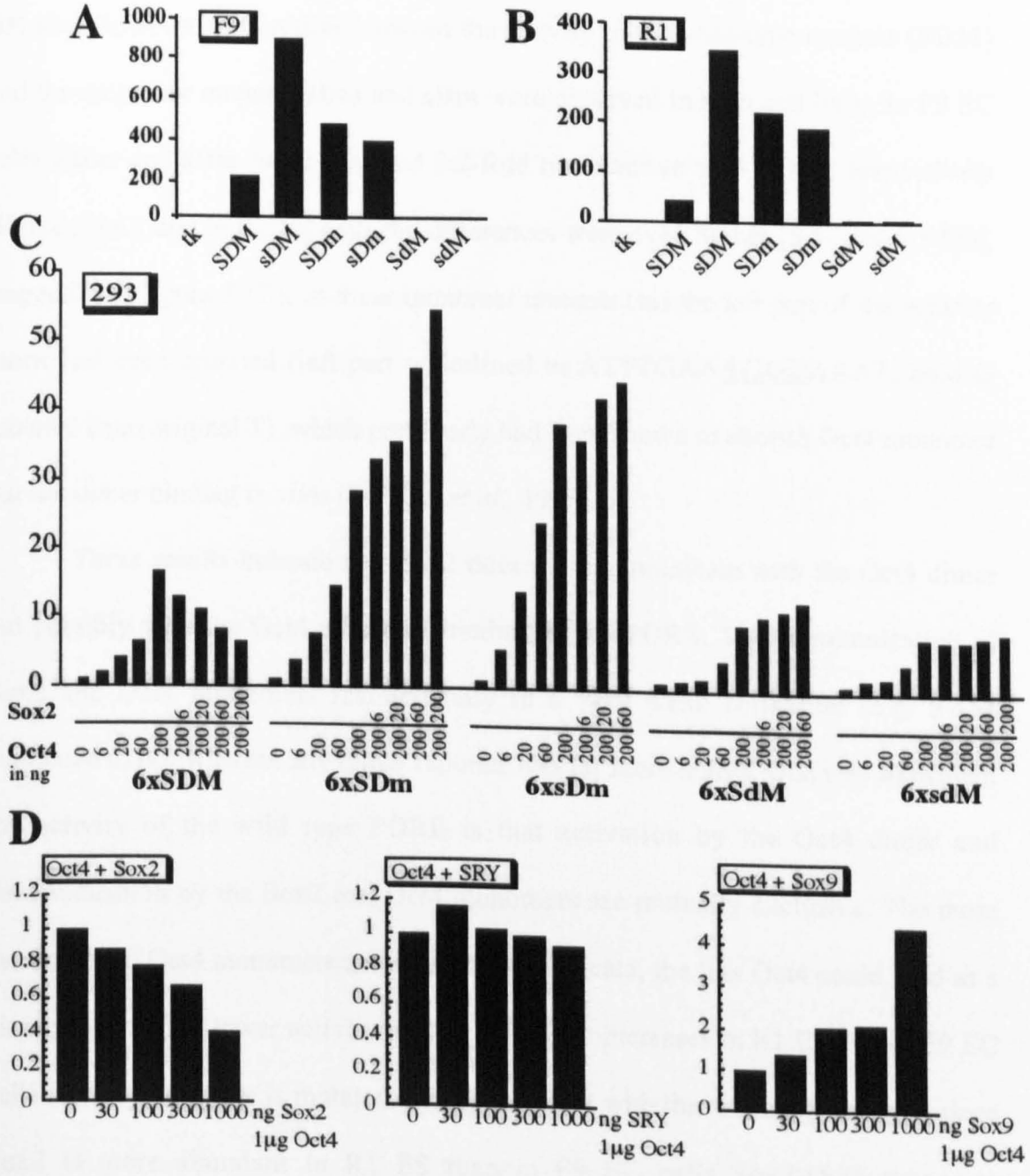
Transfection assays of *OPN* enhancer and derivatives thereof

(A, B) Transfection of 6xSDM *luciferase* reporter plasmid and derivatives with mutated Sox, Oct4 dimer or Oct4 monomer sites (s, d, m) into F9 EC (A) and R1 ES cells (B). An empty *tk-luciferase* reporter was used as a control. The results indicate that the Oct4 dimer is more active than the monomer and that the Sox element is not an independent negative element.

(C) Cotransfection of 6xSDM reporter plasmid and derivatives thereof with increasing amounts of Oct4 expression vector and increasing amounts of Sox2 expression vector with constant amounts of Oct4. This assay analyzes the effect of mutations in the recognition sequences on Oct4's activating potential and Sox2's repressing potential. Sox2 only represses transactivation in the presence of an intact PORE (DM) site.

(D) Repression is Sox2 specific. Cotransfection of 6xSDM reporter plasmid with constant amounts of Oct4 expression vector and increasing amounts of different HMG-protein expression vectors, namely Sox2, Sry and Sox9.

All values are relative luciferase activities.


A
F9
1000
800
600
400
200
0
tk
SDM
sDM
SDm
sDm
SdM
sdM
B
R1
400
300
200
100
0
C
293
60
50
40
30
20
10
0
Sox2
Oct4
in ng
0 6 20 60 200 200 200 200 200
6xSDM
6xSDm
6xsDm
6xSdM
6xsdM
D
Oct4 + Sox2
1.2
1
0.8
0.6
0.4
0.2
0
0 30 100 300 1000 ng Sox2
1µg Oct4
Oct4 + SRY
0 30 100 300 1000 ng SRY
1µg Oct4
Oct4 + Sox9
5
4
3
2
1
0
0 30 100 300 1000 ng Sox9
1µg Oct4

of **SDm** and **SdM**, respectively. On the contrary, Sox2 appeared to act positively as soon as the PORE was mutated.

The reporter plasmids were highly active only with an intact dimer-binding (**D**) site. However, differences between the activity of the wild-type reporter (**SDM**) and the monomer mutants **SDm** and **sDm** were observed in both cell lines. In F9 EC cells **SDm** and **sDm** were 1.8- and 2.2-fold more active than **SDM**, respectively (figure 5.1A), and in R1 ES cells the differences were even higher (5.5- and 6.5-fold, respectively; figure 5.1B). In these monomer mutants (**m**) the left part of the octamer motif had been mutated (left part underlined in ATTTGAAAGGCAAAT, bold G mutated from original T), which previously had been shown to abolish Oct4 monomer but not dimer binding *in vitro* (Botquin *et al.*, 1998).

These results indicate that Sox2 does not communicate with the Oct4 dimer but possibly with the Oct4 monomer binding to the PORE. The communication of Sox2 and Oct4 monomers resulted only in a very weak activation (e.g. **SdM** compared to tk), whereas any dimer reporter was far more active. One way to explain the activity of the wild type PORE is that activation by the Oct4 dimer and communication by the Sox2 and Oct4 monomers are mutually exclusive. The more the Sox2 and Oct4 monomers are able to communicate, the less Oct4 could bind as a dimer, resulting in lower activity of the **SDM**. The increases in R1 ES and in F9 EC cells if the octamer site is mutated are in agreement with this notion. Moreover, since Sox2 is more abundant in R1 ES than in F9 EC cells Sox2-Oct4 monomer communication might be more likely in R1 ES than in F9 EC cells.

5.3.2 Sox2-specific repression depends on intact PORE and Sox element

To directly assess the effect of Sox2 on Oct4 monomer and dimer activities, cotransfection experiments were performed in 293 cells, lacking both of these factors.

Increasing amounts of Oct4 and Sox2 expression vectors were cotransfected with wild-type and mutant reporter plasmids.

Sox2 alone did not have any effect on reporter activity. In the absence of Sox2, reporter activity increased with addition of increasing amounts of Oct4 expression vector (figure 5.1C). Cotransfecting increasing amounts of Sox2 with a constant amount of Oct4 resulted in activation (**SDm**), repression (**SDM**) or no significant change (**sdM**) in reporter activity, depending on the reporter construct tested (figure 5.1C). A negligible effect was observed when reporter plasmids with a mutant Sox site were used (**sDm** and **sdM**). Furthermore, Sox2-mediated repression of Oct4 activity is only observed when Oct4 can bind to the PORE either as a dimer or as a monomer. As soon as only one of both can bind, Sox2 increases reporter activity mediated by Oct4, albeit from different levels (**SDm** and **SdM**). In contrast, Oct4 monomer- or dimer-mediated activity remained unchanged upon the addition of Sox2, when the Sox2 site had been mutated (**sdM** and **sDm**). Therefore, activation and repression appear to be mediated at least in part by the canonical Sox site located 34 base pairs upstream of the PORE. Sox2 repression depends on an intact PORE. If the PORE is mutated the Sox element activates expression.

These results are comparable to those in F9 EC and R1 ES cells (figure 5.1A, B). In both cases, the loss of Oct4 monomer binding increased reporter activity (compare **SDM** to **SDm** and **sDm**). In addition, the monomer plasmids were stimulated less than the wild-type reporter (compare **SDM** to **SdM** and **sdM**). However, in contrast to the results obtained with the embryonal cell lines, the difference in activity was less pronounced in 293 cells.

Next, the question whether other Sox proteins also repress Oct4-mediated activity was to be answered. To this end, the effect of Sry and Sox9 on Oct4 mediated

activation of *OPN* was compared to that of Sox2. Increasing amounts of HMG-box expression vectors were cotransfected with a constant amount of Oct4 expression vector and 6xSDM reporters into 293 cells. Neither Sry nor Sox9 were able to repress Oct4-mediated transactivity (figure 5.1D). The presence of Sry did not affect transcription, while the addition of Sox9 stimulated activity of the SDM reporter. Taken together, these results point to a Sox2-specific repression mechanism of Oct4-mediated activation of the SDM element.

5.3.3 Sox2 and Oct4 form a heterodimer on the PORE

The Oct4 dimer contributes to the predominant transcriptional activity of the *OPN* PORE element. I have shown above that Sox2 acts as a repressor only when the Oct4 monomer and dimer can bind to the PORE. Thus, Sox2 must repress Oct4 dimer activity. Sox2 is known to interact with Oct4 via protein-protein interactions (Ambrosetti *et al.*, 1997, Chapter 4). One possible mechanism by which Sox2 could repress the Oct4-mediated activity is preventing Oct4 homodimerization by interacting with the Oct4 monomer on the *OPN* enhancer. The absence of the Oct4 homodimer may account for the net decrease in transcriptional activity. In such a scenario, repression would result from physical competition between Sox2/Oct4 and Oct4/Oct4 dimer complex formation.

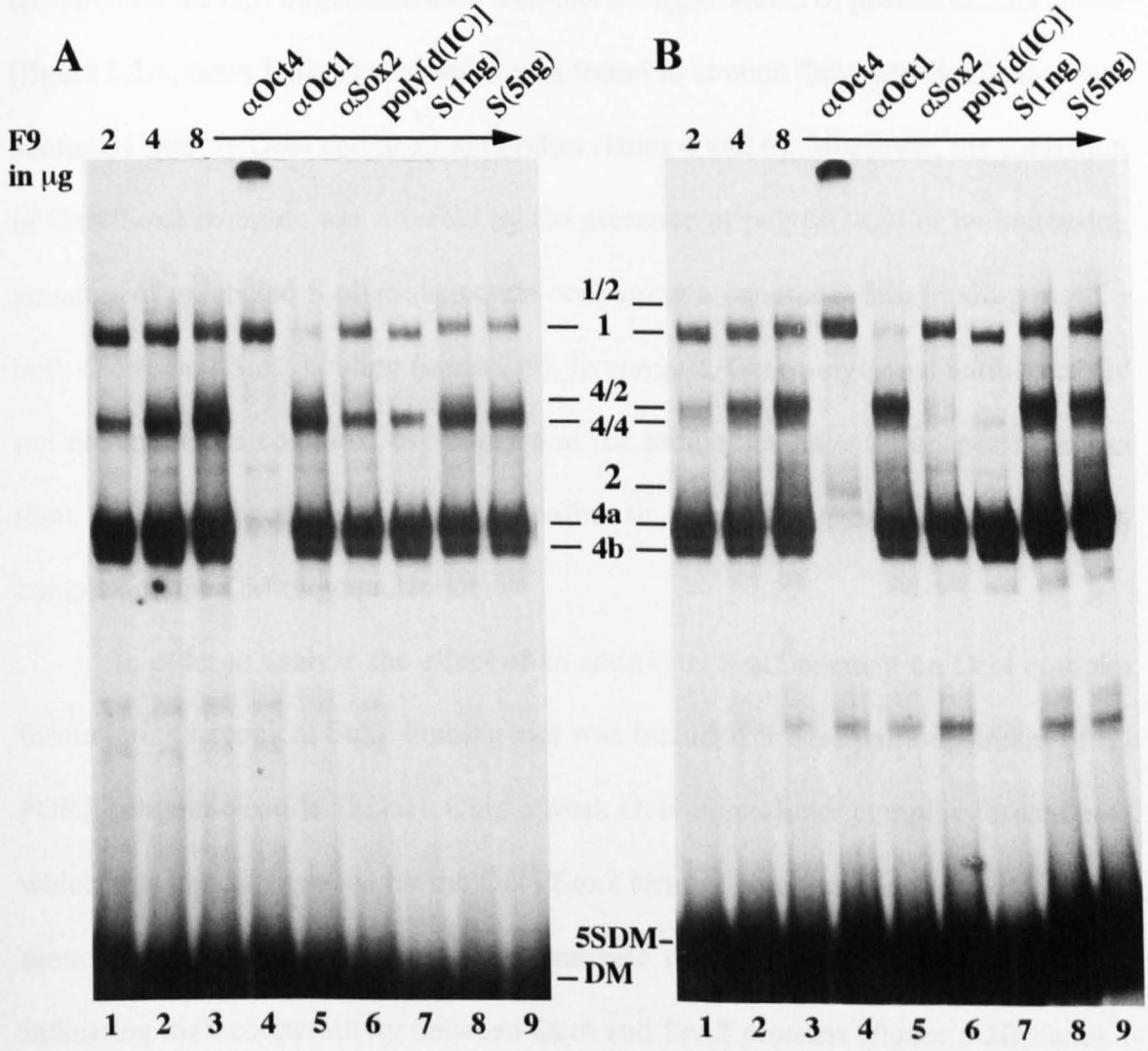
As an initial step to test this hypothesis, electrophoretic mobility shift assays (EMSAs) using the DM oligonucleotide, representing the PORE without the upstream Sox site, and F9 EC cell extracts were performed to analyze Oct4/Sox2 complex formation (referred to as O in Botquin *et al.*, 1998 and PORE in Tomilin *et al.*, 2000, Reményi *et al.*, 2001 and Chapter 3).

Increasing amounts of F9 whole cell extract were tested for Sox2 complex formation on the DM oligonucleotide (figure 5.2A). As reported previously, Oct4

Figure 5.2

F9 cell extract EMSA on PORE and 5SDM

EMSA of F9 cell extracts (2, 4 and 8μg) on (**A**) the PORE-containing DM oligonucleotide and (**B**) 5SDM where a canonical Sox site is 5 base pairs upstream of the DM sequence. Antibodies against Oct4, Oct1 and Sox2 (αOct4, αOct1, αSox2, respectively) eliminate complexes containing these proteins. Poly[d(IC)] prevents Sox2 from binding DNA and S, a Sox site containing unlabeled oligonucleotide, also competes for Sox2. The lower bands in the 5SDM EMSA are Sox2 degradation products. They do not appear at low levels of F9 cell extract and are prevented from binding DNA in the presence of poly[d(I-C)]. The protein complexes with DNA are as follows: Oct1/Sox2 (1/2); Oct1 (1); Oct4/Sox2 (4/2); Oct4/Oct4 (4/4); Sox2 (2); two differently phosphorylated Oct4 monomers (4a, 4b; V.B. unpublished).


A
αOct4
αOct1
αSox2
poly[d(IC)]
S(1ng)
S(5ng)
F9
in μg
2 4 8
1/2
1
4/2
4/4
2
4a
4b
5SDM-
DM
1 2 3 4 5 6 7 8 9
B
αOct4
αOct1
αSox2
poly[d(IC)]
S(1ng)
S(5ng)
2 4 8
1 2 3 4 5 6 7 8 9

formed monomer (4a and 4b) and homodimer (4/4) complexes on the PORE. The two monomers probably represent differently phoshorylated forms of the Oct4 protein (Botquin *et al.*, 1998). Furthermore, a monomer of Oct1 was detectable in these experiments.

A new complex (4/2) of slightly lower mobility than the Oct4 homodimer appeared on the DM oligonucleotide with increasing amounts of protein extract added (figure 5.2A, lanes 1-3). The complex was found to contain Oct4 and Sox2, as it was abolished both by Oct4 and Sox2 antibodies (lanes 4 and 6). Moreover, the formation of Oct4/Sox2 complex was affected by the presence of poly[d(I-C)] or by increasing amounts of unlabeled S oligonucleotide containing a canonical Sox binding motif – both decreasing Sox2 binding (lanes 7-9). In contrast, Oct1 polyclonal antibodies did not recognize the complex, excluding that the antibodies have an unspecific effect (lane 5). Taken together, these results confirm that Sox2 and Oct4 can form a ternary complex on the DM oligonucleotide.

In order to analyze the effect of an additional Sox2 element on Oct4 complex formation, a canonical Sox2 binding site was included 5 base pairs upstream of the PORE (oligonucleotide 5SDM). Only a weak Oct4 homodimer complex formed here, which was almost replaced by the Oct4/Sox2 ternary complex. Furthermore, a Sox2 monomer could not be shown in the presence of Oct4, even after long exposure, indicating high cooperativity between Oct4 and Sox2 proteins (figure 5.2B, lanes 1-3).

A Sox2 monomer complex (2) of weak intensity was observed following addition of Oct4 antibodies to the binding reaction (figure 5.2B, lane 4). However, the Sox2 monomer did not form on the DM oligonucleotide (figure 5.2A, lane 4), implying that Sox2 can only bind to the PORE as a heterodimer with Oct4.

An additional mobility complex containing Oct1 and Sox2 appeared above the Oct1-DNA complex (figure 5.2B, lane 3, complex 1/2). This complex was only formed on the 5SDM oligonucleotide but not on DM (cf. lanes 3 in figure 5.1A and B). The complex formation was increased following Sox2 protein release from the Oct4/Sox2 ternary complex (lane 4). The composition of all protein-DNA complexes were confirmed by their competition with specific antibodies (lanes 4-6).

Taken together, these results demonstrate that Sox2 can form a ternary complex with Oct4 on the DM oligonucleotide, representing the PORE. Moreover, a canonical Sox2 sequence located in close proximity to the PORE (5SDM) strongly enhances the formation of an Oct4/Sox2 protein complex and reduces the binding intensity of the Oct4 homodimer.

5.3.4 Sox2 binds to the non-octamer half of the PORE

The abovementioned results demonstrate the formation of an Oct4/Sox2 ternary complex on the PORE (DM oligonucleotide) (figure 5.2A). In order to assay the binding properties of the heterodimer, EMSAs of recombinant Oct4 (rOct4) and Sox2 (rSox2) proteins with DM and mutated versions thereof were performed (figure 5.3A).

The recombinant Sox2 can form a monomer (2) on **DM**, in the absence of rOct4 (figure 5.3A, lane 2). A similar rSox2 binding activity was observed on the **Dm** oligonucleotide containing a mutated octamer site (figure 5.3A, lane 8; figure 3B). In contrast, Sox2 binding activity on the **dM** oligonucleotide where the non-octamer half-site of the PORE contains a point mutation was markedly reduced (figure 5.3A, lane 12; figure 3B). These results suggest that Sox2 recognizes and binds to this part of the PORE. Comparison to a canonical Sox2 recognition element reveals that the

Figure 5.3

EMSAs on PORE and derivatives with recombinant proteins

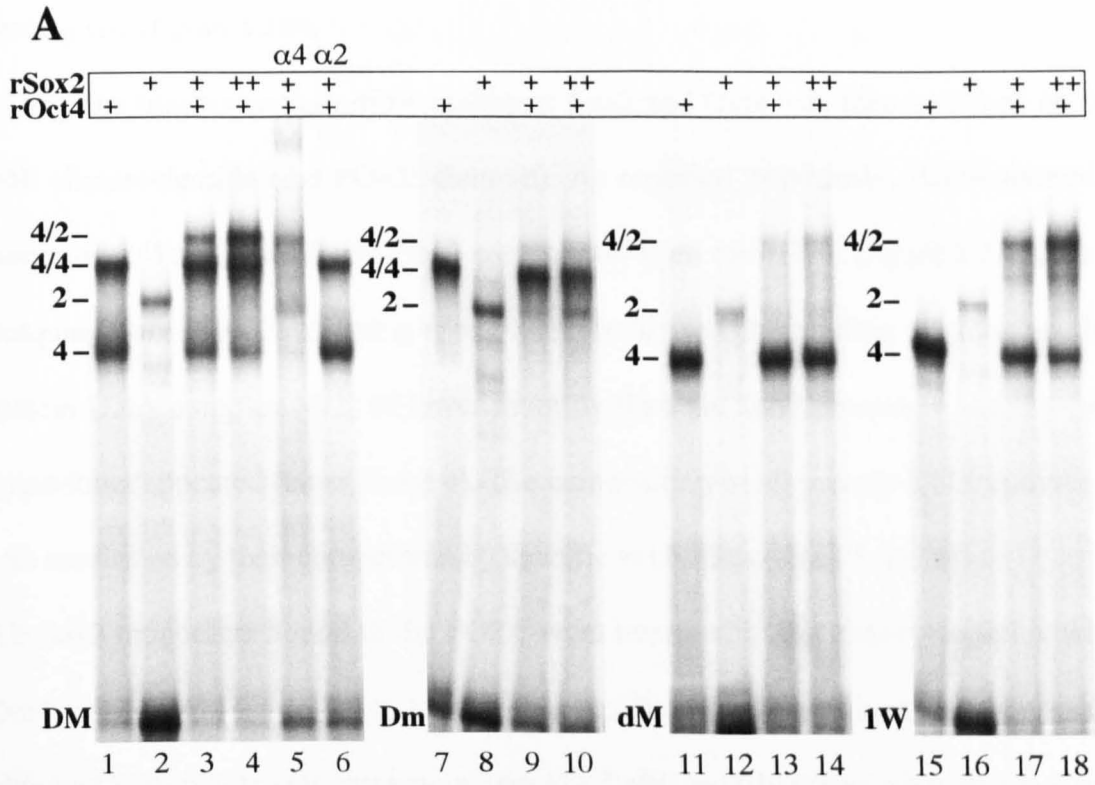
(**A**) Binding pattern of recombinant Sox2 and Oct4 (rSox2, rOct4) on the PORE (DM), its monomer mutant (Dm), dimer mutant (dM) or 1W (part of the Ig heavy chain gene enhancer element).

(**B**) Oligonucleotide sequences and their ability to bind Sox2 compared to the consensus Sox2 site.

(**C**) EMSA as in (A) on 5SDM where a canonical Sox site is located 5 base pairs upstream of the DM sequence. Increasing Sox2 levels cause the appearance of a multiprotein complex (MPC). Sox2 degradation products are detectable with Sox2 only and when antibodies against Oct4 are added to the reaction.

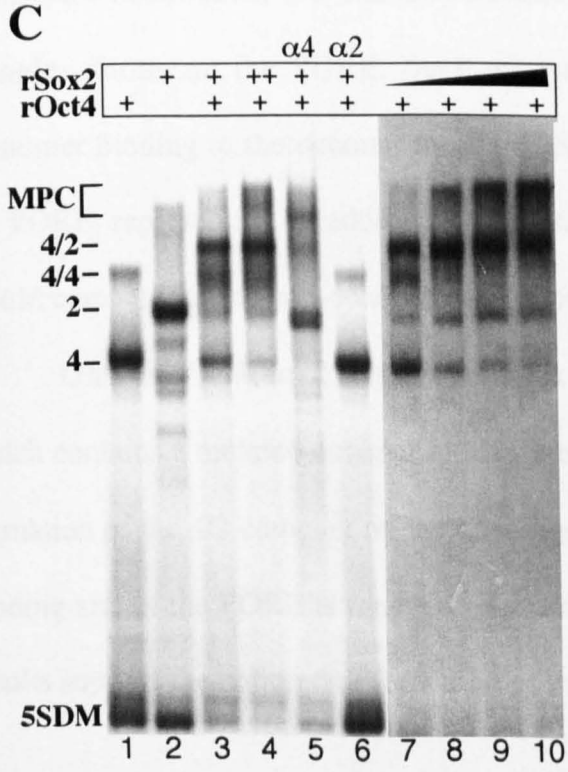
(**D**) EMSA with Oct4 and Sox2 on the phasing mutant P+1 where one nucleotide was inserted between the two half-sites of the PORE.

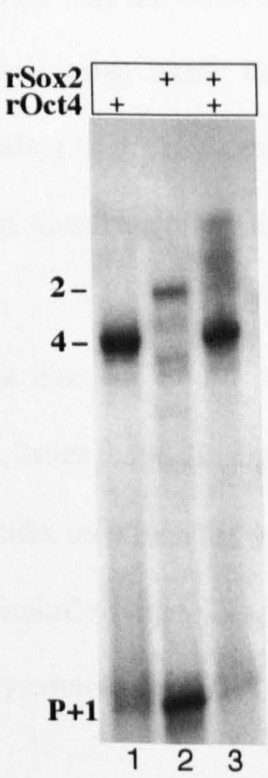
Antibodies against Oct4, and Sox2 (α4, α2) abolish complexes containing these proteins. The protein complexes with DNA are as follows: multiprotein complex consisting of Sox2 and Oct4 proteins (MPC); Oct4/Sox2 (4/2); Oct4/Oct4 (4/4); Sox2 (2); Oct4 monomer (4); lower bands represent Sox2 degradation products.



B

	Sox2/DNA complex
Sox2 consensus site **CTTTGTT**	+++
DM **ATTTGAA**ATGCAAAT	++
Dm **ATTTGAAAG**GCAAAT	++
dM **ATGTGAA**ATGCAAAT	+
1W **TTTAGAA**ATGCAAAT	+





non-octamer half-site of the PORE represents a half-core motif of a Sox2 consensus binding site (figure 5.3B).

The binding pattern of recombinant Sox2 and Oct4 was then analyzed on the **DM** oligonucleotide (the PORE element). As reported previously, rOct4 formed a monomer- (4) and homodimer-DNA complex (4/4) on the PORE (figure 5.3A, lane 1; Botquin *et al.*, 1998). Following rSox2 addition to the Oct4 binding mixture, another protein-DNA complex (4/2) of lower mobility than the Sox2 monomer and the Oct4 homodimer appeared (lanes 3 and 4). The composition of all protein-DNA complexes was confirmed by their competition by specific antibodies (lanes 5 and 6).

All Sox2 monomers bound to the PORE were involved in the ternary complex when Oct4 was added (figure 5.3A, lanes 3 and 4). This is in agreement with the results obtained with whole cell extracts, where Sox2 also readily forms a heterodimer with Oct4 on the PORE sequence (figure 5.2). As a consequence of the Sox2/Oct4 interaction, Oct4 monomer binding activity is reduced in the presence of Sox2 (figure 5.3A, lanes 1, 3 and 4). These results, as well as the fact that the Oct4 dimer and the Sox2/Oct4 heterodimer are mutually exclusive, indicate that the Oct4/Sox2 ternary complex forms on the PORE. Such an interaction is most likely through Oct4 monomer binding to the octamer motif and Sox2 binding to the non-octamer half of the PORE, representing an additional Sox binding site. Consequently, Oct4 and Sox2 would compete for the non-octamer half of the PORE.

Only an Oct4/Sox2 ternary complex of weak intensity could form on **Dm**, which contains a mutated octamer motif (figure 5.3A, lanes 9 and 10; figure 3B). The formation of the 4/2 complex on the **dM** oligonucleotide, on which the potential Sox2 binding site of the PORE is mutated, was affected similarly (lanes 13 and 14). These results support the notion that formation of the ternary complex on *OPN* requires both

intact half-sites of the PORE – an octamer site and an alternative Sox binding site. They reveal that in *OPN*, the binding of the ternary complex requires an intact PORE element on which the Oct4 homodimer and the Oct4/Sox2 heterodimer can compete for the same binding site.

The Ig heavy chain gene enhancer element (1W) contains a similar potential recognition site for Sox2 close to the octamer motif but does not contain a POU dimerization site. 1W was tested for ternary complex formation (figure 5.3B; compare TTTAGAA from 1W to ATTTGAA of the PORE), and although only a Sox2-DNA complex of weak intensity formed, a relatively strong Oct4/Sox2 ternary complex appeared on the 1W oligonucleotide (figure 5.3A, lanes 17 and 18).

Then the effect of an additional Sox2 element in close proximity to the octamer site on ternary complex binding activity was tested. In the presence of the upstream Sox2 element on 5SDM, the formation of the ternary complex was greatly enhanced and, consequently, the Oct4 monomer and dimer complexes were strongly reduced (figure 5.3C, lanes 1, 3 and 4). In addition, a multiprotein complex (MPC) became evident (lanes 3 and 4). Upon addition of the Oct4 antibody, Sox2 is released and forms a monomer of stronger intensity on the DNA (lane 5). The Sox2 antibody captures all Sox2 molecules, resulting in only the Oct4 monomer and dimer binding to DNA and the absence of an MPC (lane 6). This suggests that the MPC contains both recombinant proteins. Therefore, this MPC may represent Sox2/Sox2/Oct4, Oct4/Oct4/Sox2 or even higher-order protein-DNA complexes.

In the presence of Sox2 alone, a lower mobility complex was evident, which may consist of Sox2 homodimers or two individual Sox2 molecules bound to the two potential Sox2 binding sites on the 5SDM oligonucleotide (lane 2). In the presence of increasing amounts of Sox2 protein and constant amounts of Oct4 protein, formation

of the multiprotein complex on the 5SDM oligonucleotide was even more pronounced (lanes 7-10). Under these conditions, only weak or no binding of the Oct4 monomer and homodimer could be detected; however, the formation of the Oct4/Sox2 complex formation was unaffected.

5.3.5 Oct4/Sox2 ternary complex is cooperative

Botquin *et al.* (1998) have shown that POU proteins cooperate with each other to form dimers. A logical following question was whether Oct4 and Sox2 behave in a similar fashion. Phasing mutations, by inserting or deleting one or more nucleotides between two binding sites, were shown to abolish POU dimer formation (Botquin *et al.*, 1998). Likewise, figure 5.3D shows that Oct4 and Sox2 cannot heterodimerize on the P+1 oligonucleotide, in which one nucleotide had been inserted between the two halves of the PORE. Negligible amounts of 4/2 complex formed on P+1 (figure 5.3D, lane 3). These results suggest a cooperative interaction between Oct4 and Sox2 on the PORE.

5.3.6 Oct4/Sox2 heterodimer is more stable than Oct4 homodimer

Sox2 mediates a strong repression of Oct4 transcriptional activity (see figure 5.1). This could be attributed to a greater stability of the Oct4/Sox2 ternary complex on the PORE compared with the Oct4 homodimer. To test this hypothesis, the stability of Oct4 and Sox2-containing complexes was investigated. To this end, off-rate EMSAs using F9 extracts and the DM oligonucleotide were performed to assess the complexes' dissociation (figure 5.4A). To distinguish between Oct4 homodimer and Oct4/Sox2 heterodimer complexes, poly[d(I-C)] was added to prevent Sox2 from binding to the DNA in one experiment (compare figures 5.4A and B).

Figure 5.4

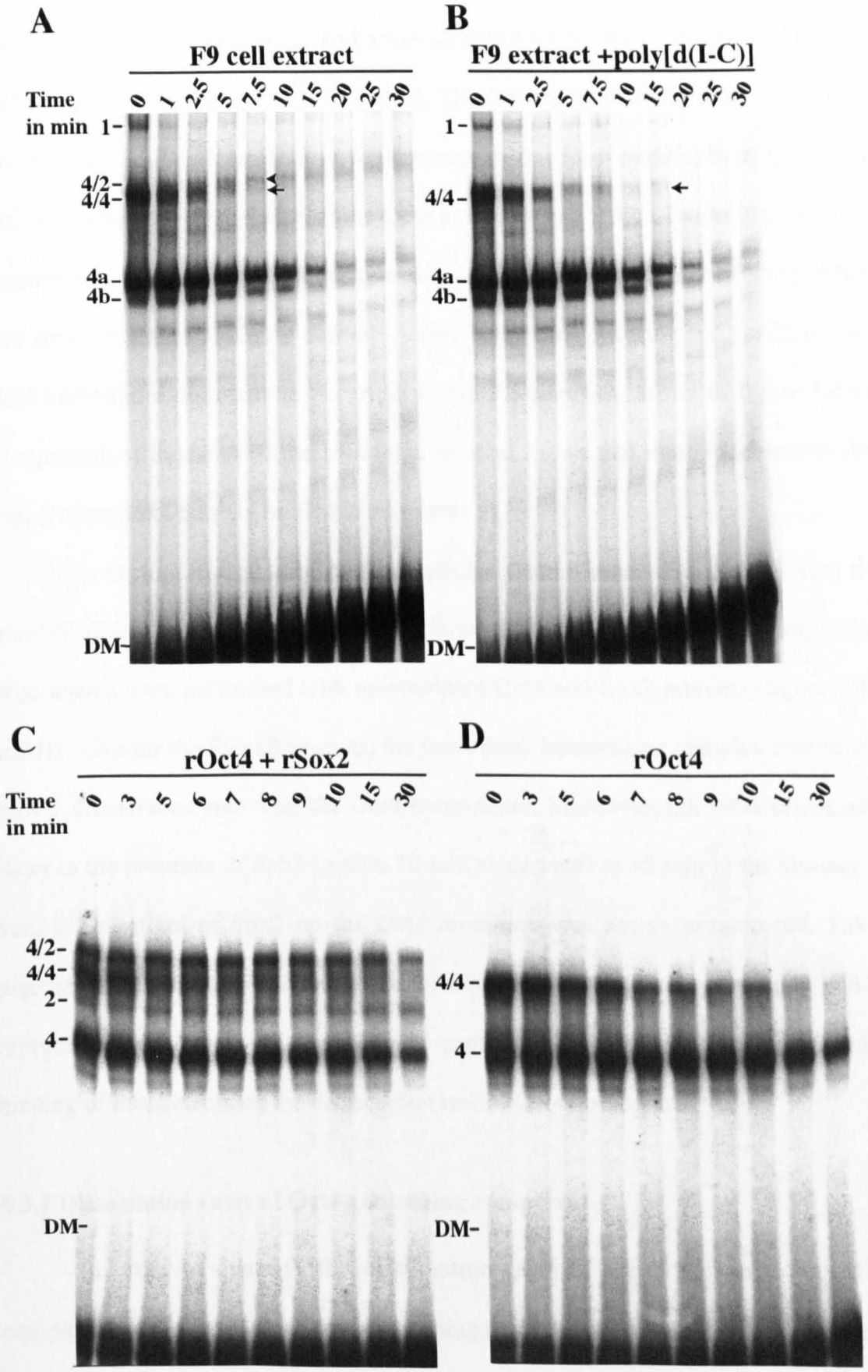
Off-rate EMSA showing stability of protein-DNA complexes

A binding reaction is prepared into which a 500-fold excess of unlabeled oligonucleotide is added after the proteins have been allowed to bind to the labeled probe. Aliquots of the reaction mixture are loaded onto a gel at regular time intervals between 1 and 30 min after addition of the unlabeled oligonucleotide. Dissociating proteins become undetectable on the gel as they re-associate to the unlabeled oligonucleotide.

(**A, B**) Stability of proteins from F9 cell extract and F9 extract with poly[d(IC)], which circumvents Sox2 binding to the PORE (DM).

(**C, D**) Off-rate experiment of recombinant Oct4 (rOct4) and recombinant Sox2 (rSox2) off DM. This excludes an effect of any factor within the whole cell extract other than Oct4 and Sox2 on the stability of the protein-DNA complexes.

The protein complexes with DNA are as follows: Oct4/Sox2 (4/2, emphasized by arrowhead); Oct4/Oct4 (4/4, emphasized by arrow); Sox2 (2); Oct4 monomer (4); two differently phosphorylated Oct4 monomers (4a, 4b; V.B. unpublished).


A
F9 cell extract
Time in min
0 1 2.5 5 7.5 10 15 20 25 30
1-
4/2-
4/4-
4a-
4b-
DM-
B
F9 extract +poly[d(I-C)]
0 1 2.5 5 7.5 10 15 20 25 30
1-
4/4-
4a-
4b-
DM-
C
rOct4 + rSox2
Time in min
0 3 5 6 7 8 9 10 15 30
4/2-
4/4-
2-
4-
DM-
D
rOct4
0 3 5 6 7 8 9 10 15 30
4/4-
4-
DM-

The Oct4/Sox2 ternary complex remained bound to DM longer than the Oct4 homodimer (compare arrowhead and arrow in figure 5.4A), which dissociated rapidly under both conditions (figures 5.4A and B). The Oct4 monomer was more stable than the Oct4 homodimer. Interestingly, the presence of Sox2 lowered the binding activity and accelerated the dissociation of the Oct4 monomer and dimer. By preventing Sox2 binding, Oct4 homodimer binding was still detectable after 15 min of competition (see arrow in figure 5.4B). In contrast, under conditions allowing for Sox2 binding, Oct4 homodimer dissociation occurred within 7.5 min (see arrow in figure 5.4A). Comparison of the lanes of the 15-minute interval shows that more monomer is still bound when Sox2 binding on DM is inhibited.

To exclude the possibility that cellular factors other than Sox2 affect the stability of the Oct4 monomer- and homodimer-DNA complexes, similar competition experiments were performed with recombinant Oct4 and Sox2 proteins (figure 5.4C and D). Like for the F9 cell extracts, the Oct4/Sox2 heterodimer complex exhibited a slower dissociation rate than the Oct4 homodimer. Moreover, the latter dissociated faster in the presence of Sox2 (within 10 min as opposed to 30 min in the absence of Sox2). The effect of Sox2 on the Oct4 monomer was not as pronounced. Taken together, the effect of Sox2 on the dissociation rate of the Oct4 homodimer can be reproduced *in vitro* using recombinant proteins. These results demonstrate that binding of Sox2 increases the dissociation rate of Oct4 homodimer.

5.3.7 Dissociation rates of Oct4-containing complexes

In order to quantify the dissociation rates of the different protein-DNA complexes, a similar competition assay using recombinant proteins was performed in an extended time course. The 5SDM oligonucleotide was used in parallel with DM to determine the influence of a Sox2 binding site on the stability of the different

complexes. The intensities of the complexes were quantified using the Phosphorimager (figure 5.5).

Percentage of bound complex at a specific time (percentage of [bound protein-DNA complex at time x]/[$\sum$ of total bound protein-DNA complex at time x]) was then plotted as a function of time on a semi-logarithm graph (figures 5.5C, D). The half-lives of the protein-DNA complexes in minutes were interpolated from the graphs (figure 5.5C-E) and the dissociation rate constants (k_d) were calculated from the equation $\ln 2/T_{[50\% \text{ complex bound}]}$ (figure 5.5E). The half-life of the Oct4/Sox2 heterodimer on the PORE is 5 times longer than that of the Oct4 homodimer. Moreover, the presence of a Sox2 recognition element upstream of the PORE further destabilizes the Oct4 monomer and dimer complexes. The half-life of the heterodimer on 5SDM is 11-fold longer than that of the homodimer. This clearly proves that the Sox2/Oct4 heterodimer complex is more stable than the homodimer.

In contrast to all Oct4-containing complexes, binding of the Sox2 monomer complex to both oligonucleotides increases during the time course of the experiment. This indicates that the Sox2 protein-DNA complex by itself is remarkably stable and as such favors the formation of the newly Oct4/Sox2 ternary complex. The dissociation of the Oct4/Sox2 complex reflects only the dissociation of the Oct4 molecule from the DNA. The Sox2 molecule remains bound, resulting in an increase in Sox2 monomer bound to DNA over time.

5.3.8 Upstream Sox binding site recruits Sox2 to the PORE

The transfection results in figure 5.1 indicate that the Sox2 site located 34 base pairs upstream of the PORE is necessary for repression. In addition, the EMSAs show that Sox2 binds to the PORE (DM) in a stable Oct4/Sox2 heterodimer complex. The Sox2 element presumably acts by recruiting Sox2 proteins, which in turn contribute to

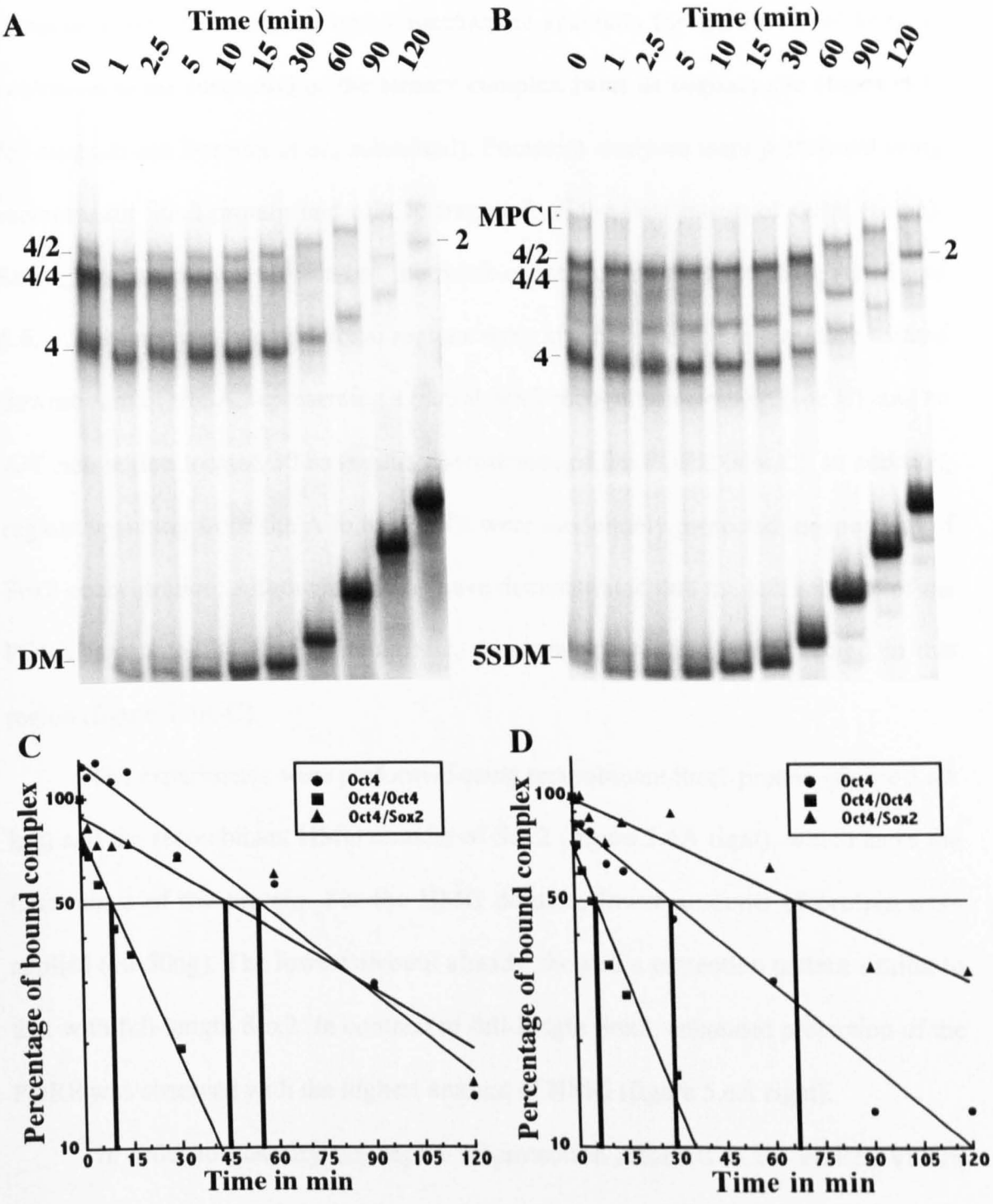
Figure 5.5

Dissociation constants of protein-DNA complexes

(**A,B**) Off-rate experiments of recombinant Oct4 and Sox2 of DM (**A**) and 5SDM (**B**) as in figure 5.4. The protein complexes with DNA are as follows: multiprotein complex consisting of Sox2 and Oct4 proteins (MPC); Oct4/Sox2 (4/2); Oct4/Oct4 (4/4); Sox2 (2); Oct4 monomer (4).

(**C,D**) Plot of percentage of bound complex at a specific time (percentage of [bound protein-DNA complex at Time-x]/[$\sum$ of total bound protein-DNA complex at Time-x]) on semi-logarithm graph for DM (**C**) and 5SDM (**D**).

(**E**) The half-lives of the protein-DNA complexes in minutes were interpolated from the graph. Dissociation rate constants (k_d) in s^{-1} were calculated with the equation $\ln 2/T_{[50\% \text{ complex bound}]}$.



E

DNA	DM			5SDM		
protein	4/4	4	4/2	4/4	4	4/2
half life	9'	55'	45'	6'	30'	67'
kd	0.0013	0.00021	0.00026	0.0016	0.0003	0.00017

the formation of a ternary complex on the PORE.

My collaborator Valérie Botquin (EMBL, Heidelberg) conducted DNase I footprint assays to determine which mechanism accounts for the ability of Sox2 to contribute to the formation of the ternary complex from its cognate site (figure 5.6; for methods see Botquin *et al.*, submitted). Footprint analyses were performed using recombinant Sox2 protein and a PCR fragment of the first intron of *OPN* (i-opn). Strong protection was afforded to the Sox2 consensus binding site (site A) (figure 5.6A-C). Moreover, two additional regions were highly protected: a sequence located downstream of site A, representing a partial Sox2 recognition element (site B), and an A/T rich region located 30 base pairs downstream of the PORE (site C). In addition, regions spanning from site A to the PORE were moderately protected independent of Sox2 concentration. Although EMSAs have demonstrated that the left half-site of the PORE is a potential Sox2 binding site, no strong protection was detected in this region (figure 5.6A-C).

The experiments were performed using recombinant Sox2 protein (figure 5.6A left) and the recombinant HMG domain of Sox2 (figure 5.6A right), which lacks the C-terminus of the protein. For the HMG domain, lower amounts of protein were applied (10-50ng). The lowest amount already showed a protection pattern similar to that with full-length Sox2. In contrast to full-length Sox2, enhanced protection of the PORE was observed with the highest amount of HMG (figure 5.6A right).

In order to quantify the degree of protection afforded to the PORE, Valérie Botquin carried out phosphorimager analyses of the full-length Sox2 DNase I footprint (figure 5.6B). A comparison of the relative band intensities between DNase I digested DNA in the presence (+ Sox2; red) and absence (- Sox2; black) of Sox2 was

Figure 5.6

Footprint assays determining location of Sox2 binding on DNA

(**A**) Footprint of increasing amounts of recombinant Sox2 (rSox2) and the HMG domain (rHMG) of Sox2 on a DNA fragment of the first intron of *OPN*. Sites A, B, C and the PORE were bound by Sox2 or HMG and thus protected from DNase I.

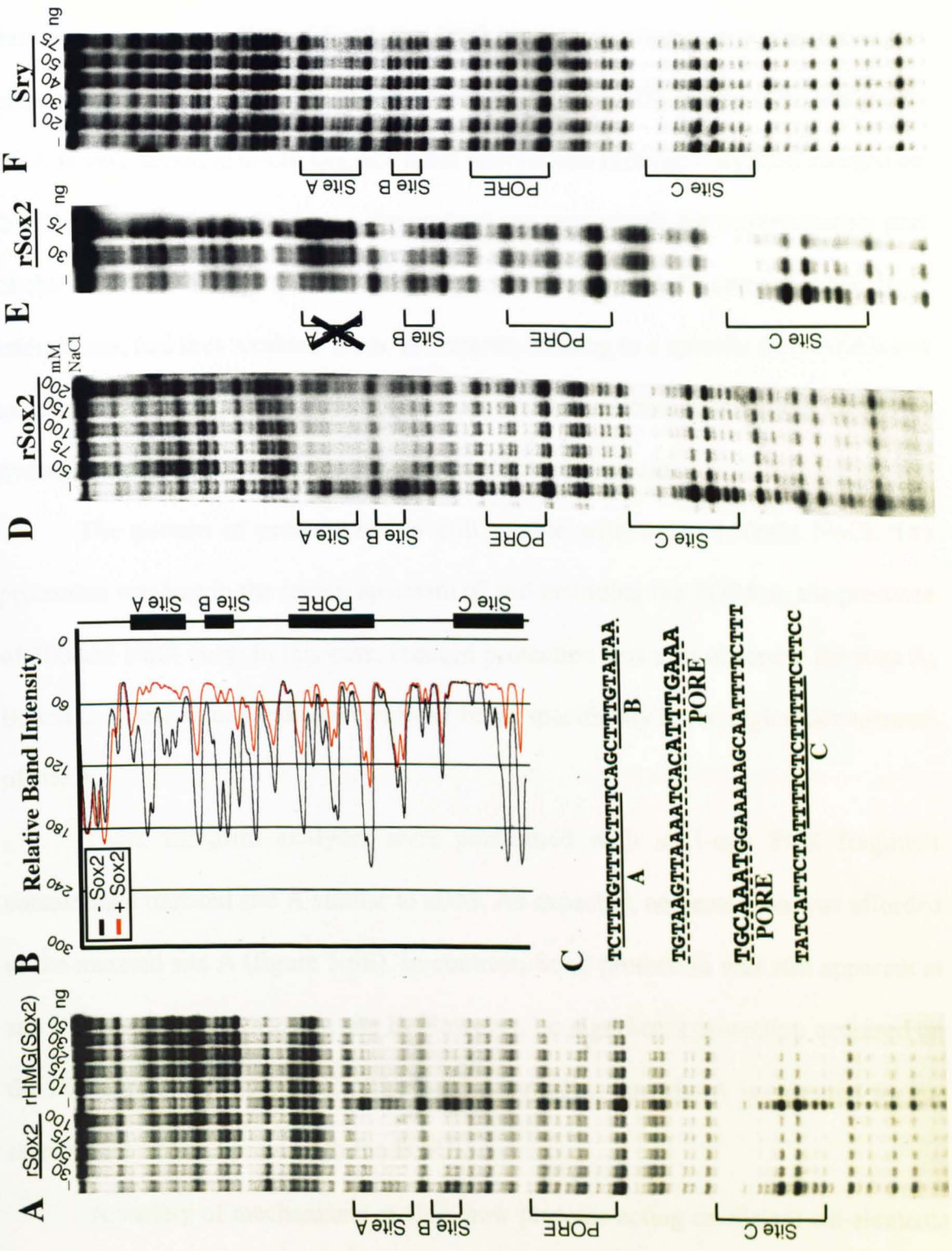
(**B**) Relative band intensities without (black) and with (red) Sox2 protection, measured by a phosphoimager. Lower intensities as shown with Sox2 at sites A, B, C and the PORE indicate protection from DNase I due to the binding of Sox2 to these areas.

(**C**) Sequence of the DNA fragment of the first intron of *OPN* displaying the sites protected by Sox2.

(**D**) Footprint of Sox with increasing amounts of NaCl proving that Sox2 binds specifically to this region.

(**E**) Footprint with increasing amounts of rSox2 on a derivative of the DNA fragment with a mutated site A. Protection of site A, the PORE and sequences in between is not seen anymore. Only site B is protected slightly, but less than that of the wild-type DNA (figure 5.6A). Site C is unaffected by the mutation.

(**F**) Increasing amounts of Sry confer no protection in footprint assay, revealing that binding to all sites is Sox2 specific.


A
rSox2
rHMG(Sox2)
ng
Site A
Site B
PORE
Site C
B
Relative Band Intensity
- Sox2
+ Sox2
C
TCTTTGTTCTTTCAGCTTTGTATAA
A
B
TGTAAGTTAAAATCACATTTGAAA
PORE
TGCAAATGGAAAAAGCAATTTTCTTT
PORE
TATCATTCTATTTTCTCTTTTTCTTCC
C
D
rSox2
mM NaCl
E
rSox2
F
Sry

obtained. In the presence of Sox2, the band intensities measured on the PORE region were significantly reduced to about 45%.

To discount the possibility that the protection between site A and the PORE is due to nonspecific binding of Sox2, the Sox2 binding reaction was performed in the presence of increasing salt (NaCl) concentration (figure 5.6D). Nonspecific binding to DNA is very sensitive to salt because ionic interactions (between negative charges on the DNA and positive charges on the protein) are responsible for a considerable part of the favorable energy that drives nonspecific binding. NaCl competes for these interactions, and thus weakens them. In contrast, binding to a specific DNA site is not as sensitive to NaCl, as this interaction includes the additional stabilizing energy of hydrogen bonds between the protein and the DNA binding site.

The pattern of protection was still visible with up to 150mM NaCl. The protection was lost in the region upstream of and including the PORE in the presence of 200mM NaCl only. In this case, reduced protection was also observed for sites A, B, and C. These results indicate that Sox2 binds specifically to the region downstream of site A.

Then, footprint analyses were performed with an i-opn PCR fragment containing a mutated site A similar to **sDM**. As expected, no protection was afforded to the mutated site A (figure 5.6E). In contrast, Sox2 protection was still apparent at site C and to a lesser extent at site B. However, no significant protection occurred on the PORE region (figure 5.6E). These results suggest that site A is essential for the recruitment of Sox2 proteins to the PORE.

A variety of mechanisms explain how proteins acting on distant cis-elements may regulate gene expression, including looping, twisting, sliding and oozing (reviewed in Ptashne, 1986). The mechanism of oozing describes the binding of a

regulator protein to its cognate sequence as helping the binding of another molecule to adjacent sequences. This in turn helps another protein bind next to it, and so on, until a procession of proteins has oozed out from the control sequence to the target element. Sliding involves the binding of a protein to a DNA element with high binding affinity, from which it slides along the DNA where binding is still stable but not sufficient to allow binding initiation. For both models, a mutation of the upstream Sox site (site A) that prevents initial Sox2 binding would result in the inhibition of Sox2 heterodimerization with Oct4 on the downstream PORE.

Cotransfection experiments have shown that neither Sry nor Sox9 proteins were able to repress Oct4 activity, suggesting a Sox2-specific repression mechanism (figure 5.1D). This could be attributed to Sox2-specific sliding or oozing. To test this hypothesis, increasing amounts of recombinant Sry protein were assessed in a DNase I footprint assay with the wild-type i-opn fragment (figure 5.6F). Although binding of Sry on the *OPN* fragment occurred in the mobility shift assay, footprint analyses detected only modest protection of sites A, B and C. However, no protection appeared on the regions next to site B and on the PORE, even with the highest amount of Sry protein. This suggests that weak binding of Sry to the *OPN* fragment is not sufficient to allow sliding or oozing to occur to adjacent sequences, or that sliding or oozing is Sox2 specific.

5.3.9 Influence of site A on gene repression

The experiments described above indicated that the upstream Sox site recruits Sox2 to the PORE by sliding or oozing. Mutating the Sox site (site A) prevents the recruitment, suggesting that it is essential for the formation of a multi-Sox2 complex that reaches the PORE site. Mobility shift assays have revealed the formation of

multiprotein complexes (MPC) only when a Sox2 consensus binding site and the PORE are present in an oligonucleotide (figure 5.3C).

The next step was to determine whether this MPC formation is controlled by site A. To this end, an EMSA was performed to assess the ability of the MPC to form on the wild-type **SDM** oligonucleotide compared to one with a mutated Sox site (**sDM**).

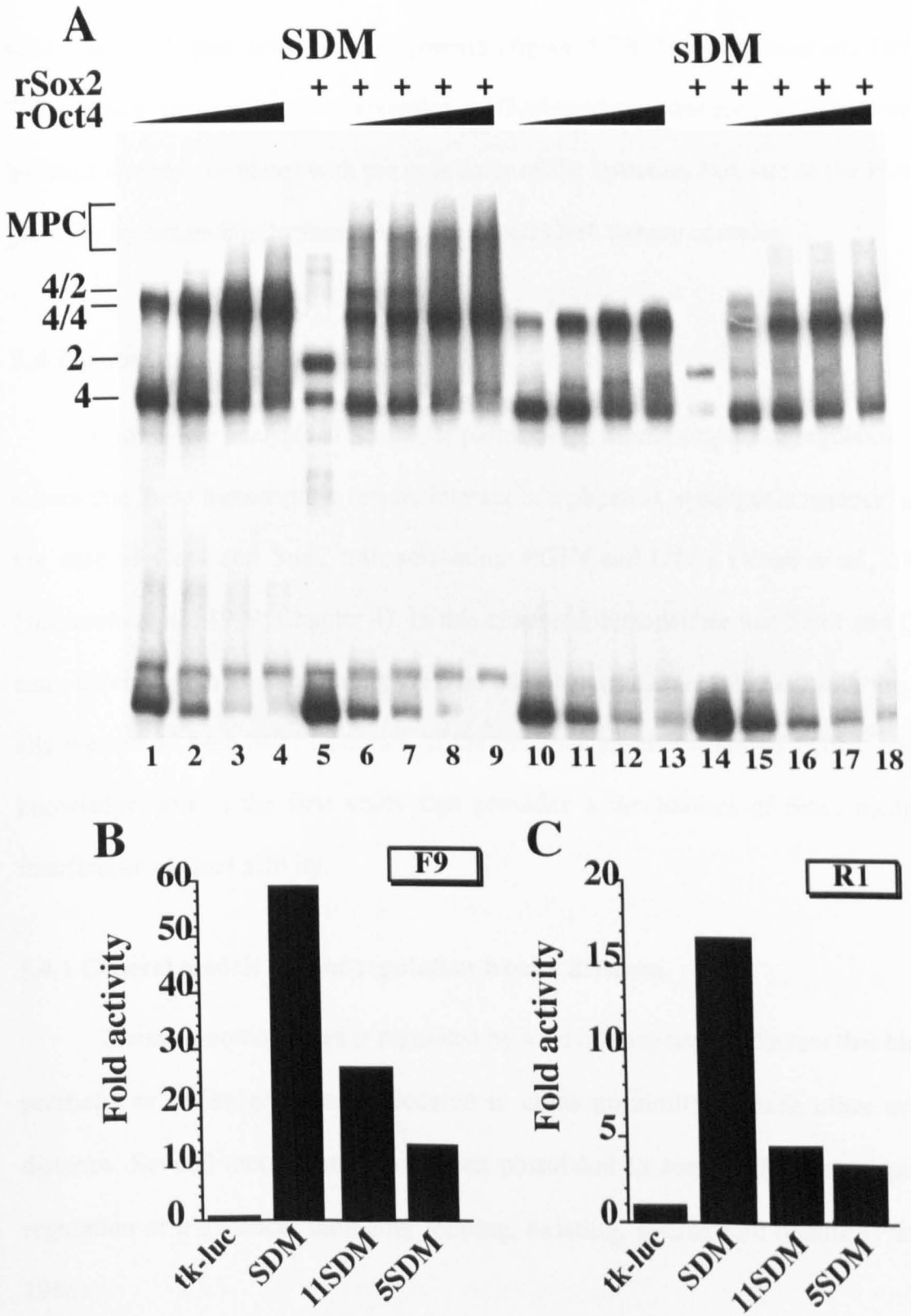
Increasing amounts of recombinant Oct4 were used in combination with constant amounts of recombinant Sox2. As expected, a multiprotein-DNA complex was detected on the **SDM** oligonucleotide (figure 5.7A, lanes 6-9). Both, Oct4- and Sox2-specific antibodies interfered with this complex, indicating that both proteins are in the MPC (cf. figure 5.3C). The complex also migrates at a higher position than that on 5SDM, indicating that it contains more protein molecules. In contrast, no higher-order Oct4/Sox2-DNA complex could be formed on the **sDM** oligonucleotide (figure 5.7A, lanes 15-18). These results indicate that the Sox2 recognition element is necessary for multiprotein complex formation.

To further demonstrate that Sox2 reaches the PORE by sliding or oozing from its recognition site, I surmised that a shorter region between the upstream Sox site and the PORE could enhance Oct4/Sox2 heterodimerization on the PORE due to more efficient sliding or oozing over a shorter stretch of DNA. For this purpose, *tk luciferase* reporter plasmids containing oligonucleotide hexamers in which the upstream Sox site was brought 5 base pairs (6x5SDM) and 11 base pairs (6x11SDM) away to the PORE were transiently transfected into F9 EC and R1 ES cell lines. Their luciferase activity was compared to that of the wild-type vector 6xSDM, in which the Sox site is 34 base pairs upstream of the PORE (figure 5.7B, C).

Figure 5.7

Oct4/Sox2 MPC formation depends on site A

(**A**) EMSA of increasing amounts of recombinant Oct4 with or without constant amounts of recombinant Sox2 on the wild- type *OPN* fragment (SDM) or a derivative thereof containing a mutated Sox site (sDM). The multiprotein complex (MPC) only binds in the presence of an intact Sox site (SDM). The other protein complexes with DNA are: Oct4/Sox2 (4/2); Oct4/Oct4 (4/4); Sox2 (2); Oct4 monomer (4). Sox2 degradation products are detectable in the reaction containing Sox2 without Oct4.
(**B, C**) Reducing the distance between Sox2 and the PORE motifs increases gene repression. Transfection assay of SDM reporter plasmid and its derivatives with only 11 (11SDM) or 5 base pairs (5SDM) between the PORE and Sox sites into F9 EC (**B**) and R1 ES cells (**C**). The -37*tkluc* reporter plasmid was used as a control.


A
SDM
sDM
rSox2
rOct4
+ + + + +
+ + + + +
MPC
4/2
4/4
2
4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
B
F9
Fold activity
60
50
40
30
20
10
0
tk-luc
SDM
11SDM
5SDM
C
R1
Fold activity
20
15
10
5
0
tk-luc
SDM
11SDM
5SDM

In both cell lines, the highest activity was obtained with the wild-type 6xSDM reporter plasmid and activation decreased the closer the Sox site was brought to the PORE (figure 5.7B, C). This repression effect was even more increased in R1 ES cell, which have a higher level of Sox2 protein (figure 5.7B, C; Botquin *et al.*, 1998). These results demonstrate that repression of Oct4-mediated transcriptional activation by Sox2 directly correlates with the proximity of the upstream Sox site to the PORE, probably by enhancing the formation of the Sox2/Oct4 ternary complex.

5.4 Discussion

Cumulative analysis of Sox-POU partnerships in transcriptional regulation has shown that these transcription factors interact in a physical, synergistic manner, as in the case of Oct4 and Sox2 transactivating *FGF4* and *UTF1* (Yuan *et al.*, 1995; Nishimoto *et al.*, 1999; Chapter 4). In this chapter I demonstrate that Sox2 and Oct4 can interact together in regulating *OPN* in an apparent distant manner, with the Sox site located 34 base pairs upstream of the octamer motif-containing PORE. To my knowledge, this is the first study that provides a mechanism of Sox2-mediated interference of Oct4 activity.

5.4.1 General models of gene regulation from a distance

Transcription of genes is regulated by a set of transcription factors that bind to promoter or enhancer elements located in close proximity to each other or at a distance. Several mechanisms have been postulated to account for transcriptional regulation at a distance, including looping, twisting, sliding and oozing (Ptashne, 1986).

The first indication of looping came from studies of gene regulation in *Escherichia coli*. Irani *et al.* (1983) showed that two operators separated by about 90 base pairs are required for efficient repression of the *gal* operon, and they suggested that interaction between the Gal repressors bound to this operator could only be possible by a mechanism of looping (Gal repressors bound to this operator as one possible mechanism of looping. In fact, they reasoned that only DNA looping could allow for this type of interaction.). This conclusion followed the observation that efficient repression was maintained when an integral number of helical turns of DNA was added or deleted between the two transcription factor binding sites, but not when the spacing was varied by half integral numbers of turns. Irani *et al.* (1983) reasoned that proteins separated by a distance can interact by simple looping only if located on the same side of the helix. This concept has been extended to other regulatory elements, like the early gene promoter of simian virus (SV40) and ß-globin gene regulation via locus control regions (LCRs) (Takahashi *et al.*, 1986; Bulger and Groudine, 1999). In addition to phasing evidence, DNase I fooprint analyses have shown that as the proteins bind, the DNA between the sites becomes alternately sensitive and resistant to DNase I cleavage at half turn intervals (Hochschild and Ptashne, 1986), since the bent DNA becomes differently exposed to DNase I.

Twisting has been proposed as a model in which conformational unwinding of DNA, following enzymatic activity of a binding protein, allows binding of transcription factors and thus regulates transcription. This model finds an example in the matrix attachment regions (MARs). MARs are DNA regions spanning 250 base pairs to several kilobases in length; they are enriched in AT nucleotides and bind to nuclear matrix. They often contain consensus topoisomerase II cleavage sites, single-stranded, kinked and curved DNA and potential binding elements for homeobox-

containing DNA-binding proteins (Boulikas, 1993). MAR elements are capable of unwinding DNA and are thought to play an important role in transcriptional activation (Bode *et al.*, 1992). Footprinting studies demonstrated that DNA unwinding often presents an aberrant pattern of sensitivity to DNAse I caused by an asymmetric distortion of DNA (Basak and Nagaraja, 1998).

Sliding is a mechanism in which a protein recognizes a specific site on the DNA and then moves along the DNA to another specific sequence where it regulates transcription, possibly by interacting with another protein. Sliding has been described for the bacteriophage T4 gene 45 protein (gp45), which serves as a sliding clamp of viral DNA replication and as the activator of T4 late gene transcription (Tinker-Kulberg *et al.*, 1996). The proliferating cell nuclear antigen (PCNA), the eukaryotic functional homolog of gp45, tethers DNA polymerases onto DNA templates in order to increase the speed and the processivity of DNA replication (reviewed in Kelman and O'Donnell, 1995). These proteins have a central channel large enough to accommodate a DNA duplex, with positive charges lining the inside of the channel and negative charges distributed along the outside (Kong *et al.*, 1992; Krishna *et al.*, 1994), clearly implying their close association with DNA. In footprint analyses, the dynamic of DNA-tracking proteins can be characterized by their extreme lability to bind a specific recognition element.

Finally, oozing is a model characterized by the binding of a transcription factor to its recognition site and aiding another molecule to bind to adjacent sequences, which in turn aids another molecule to bind next to it, and so on, until a procession of proteins has oozed out from the initial binding site to the target element (Ptashne, 1986). To date, there are no published studies of this mechanism of transcriptional regulation.

5.4.2 Possible mechanisms of Sox2 recruitment to the PORE

DNAse I footprint analyses showed the ability of Sox2 to migrate from its recognition site to the PORE (figure 5.6). Strong Sox2 protections were found in three different areas: the Sox2 consensus binding site (site A); a sequence downstream of site A, representing a partial Sox2 recognition element (site B), and an AT rich region 30 base pairs downstream of the PORE (site C). In addition, the region between site A and the PORE was moderately but specifically protected even in the presence of the lowest amount of Sox2 protein. This low protection area was probably due to a low binding affinity of Sox2 to these regions. Moreover, mutating site A abrogated the ability of Sox2 to migrate out to the PORE, indicating the potential of site A as a loading site for Sox2 proteins.

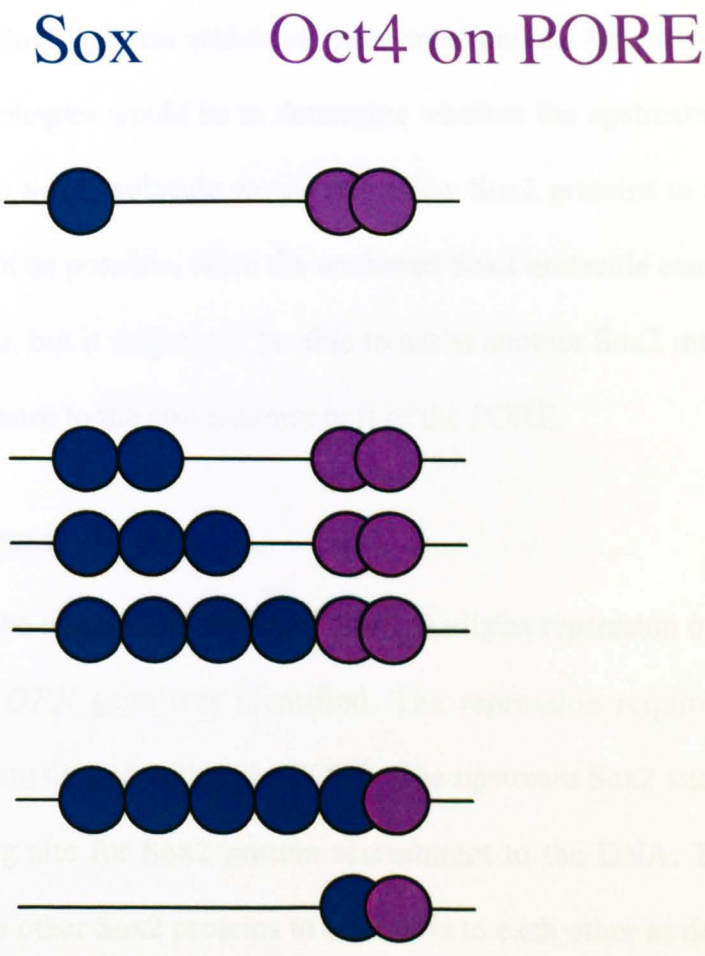
In EMSAs, Sox2 multiprotein complex formation was also dependent on the Sox consensus-binding site (Site A), confirming the data obtained in DNAse I footprint analyses (figure 5.7A). Reducing the distance between site A and the PORE was shown to directly correlate with a decrease of the *OPN* enhancer activity (figure 5.7B, C).

Footprinting assays of the oozing mechanism should result in a uniform, overall protection pattern of the DNA. The degree of protection may constantly decrease with increasing distance from the initial binding site due to incomplete traveling of the proteins along the DNA. According to earlier reports of sliding (section 5.4.1) protection of footprinting assays is less pronounced due to the dynamic of the DNA tracking protein. In twisting and looping however, patterns of protection are interrupted. The pattern of the DNase footprints presented in figure 5.6 supports the notion that Sox2 reaches the PORE by the mechanism of sliding or oozing as shown in figure 5.8. The oozing mechanism is supported by the fact that the

Figure 5.8

Mechanistic model applicable to sliding or oozing

A Sox2 molecule binds to the consensus Sox binding site upstream of the PORE. In the case of oozing this would then enhance another Sox2 molecule to bind next to it, etc. In the case of sliding the model would look similar, but the Sox2 molecule would slide along the DNA towards the PORE making the consensus Sox site available for another molecule to bind and slide downstream, etc. For both mechanisms, finally, one Sox2 molecule binds to the non-octamer site of the PORE, which was made available by the less stable Oct4 dimer.


Sox
Oct4 on PORE

protection along the DNA is fairly strong and not labile. Furthermore, proteins that slide along the DNA have so far only been reported as those which have a central channel that accommodates the DNA duplex and, to my knowledge, they have not been reported to bind as MPCs to DNA. Nonetheless, it is very difficult to distinguish between the mechanisms of sliding and oozing in biochemical experiments. One experiment that might shed light on which of these mechanisms best explains the results presented in this chapter would be to determine whether the upstream Sox site (site A) cross-linked to a Sox2 molecule would still allow Sox2 proteins to reach the PORE. Sliding would not be possible, since the anchored Sox2 molecule cannot move along the DNA anymore, but it might still be able to assist another Sox2 molecule to bind next to it and thus ooze to the non-octamer part of the PORE.

5.4.3 Repression mechanism by Sox2

In this chapter, the mechanism by which Sox2 mediates repression of the Oct4 dimer activity on the *OPN* gene was identified. The repression requires a Sox2 binding element upstream of, and within the PORE. The upstream Sox2 site serves as an initial stable loading site for Sox2 protein recruitment to the DNA. This initial loading site then allows other Sox2 proteins to bind next to each other as described in the oozing model, or might allow Sox2 to stably anchor to the DNA and slide along the DNA until it reaches the PORE. The upstream Sox site is necessary for the initiation of multiprotein complex formation. On the PORE, Sox2 cooperatively binds to Oct4 forming a ternary complex with the DNA. The Oct4 homodimer and the Sox2/Oct4 ternary complex bind to exactly the same binding site. Binding of the ternary complex is favored over the homodimer as the former is more stable. Repression is indirect, since Sox2 forms a ternary complex with Oct4 thereby preventing physical access of the activating Oct4 homodimer to the PORE.

Furthermore, repression is Sox2 specific. Neither Sry nor Sox9 repress transcription of a downstream reporter in the context of the *OPN* sequence. Functional cooperation of Sox and POU proteins is not haphazard, as particular Sox and POU factors form partnerships. For instance, on the *FGF4* enhancer, Sox10 functions best with Tst-1/Oct6/SCIP and Sox11 with Brn-1 and Brn-2 depending on the location of *FGF4* expression (Kuhlbrodt *et al.*, 1998b).

5.4.4 Sox2/Oct4 interaction on a submolecular level

Sox2 recognizes the non-octamer site of the PORE while Oct4 binds to the octamer site. The binding requirements of the ternary complex on the PORE are different than those on the *FGF4* enhancer element. On the *OPN* intron, both respective binding sites were closer to each other (compare sequence *FGF4* TCTTTGTTTGGATGCTAAT (Yuan *et al.*, 1995) and *OPN* PORE ATTTGAA ATGCAAAT). The spacing between Oct4 and Sox2 binding sites on the PORE resembles that of *UTF1* (CATTGTT ATGCTAGT) (Nishimoto *et al.*, 1999). Like on the *FGF4* enhancer element, the assembly of the ternary complex on the PORE depends on a specific spatial arrangement of both binding sites. In the case of *OPN*, the insertion of one nucleotide between both binding sites interfered with the formation of the ternary complex. In Chapter 4 of this thesis I described how Sox2 interacts with Oct4 through two separate surface patches, whereas the same surface patch of Oct4 is involved in the interaction. The sequence analysis suggests that Sox2 and Oct4 interact in a *UTF1*-like manner on the PORE, but this still remains to be shown.

In conclusion, this chapter presents evidence of a new mechanism of Sox2-mediated transcriptional regulation. The results reveal that Sox2 reaches the PORE by

migrating along the PORE through a mechanism of sliding or oozing and forms a remarkably stable ternary complex with Oct4. This prevents Oct4 dimer formation on the PORE and consequent *OPN* transcriptional activation. These findings may be used to identify other target genes regulated by Sox-POU partnerships, particularly those that harbor a Sox site at a distance from the octamer motif.

5.4.5 Addendum

In order to clarify some interpretations of the results in this chapter the following section is added to the discussion.

A misunderstanding might arise concerning the role of the Sox2 molecule at its binding site 34bp upstream of the PORE as well as one the PORE. First, I want to discuss activation mediated by the PORE (DM) and mutations thereof, in the next paragraph I will discuss these results in relationship to protein binding. In a set of experiments, the Oct4 monomer can bind and activate reporter gene expression via the upstream PORE. This was shown by cotransfection experiments with Oct4 expression vector and reporter plasmid with **dM** driving the luciferase gene (figure 5.1 and Botquin *et al.*, 1998). The Oct4 dimer activates transcription much more effectively (cf. **sDm** and **sdM** in figure 5.1C). The Oct4/Sox2 heterodimer also activates transcription, but not as efficiently as the Oct4 homodimer (Yuan *et al.*, 1995; Nishimoto *et al.*, 1999). Thus, the Sox2 itself is not a repressor, but its presence results in a net reduction in transcriptional activity if Sox2 is allowed to bind the PORE, since the Oct4/Sox2 heterodimer competes for the same binding site as the more potent Oct4 homodimer. The activating effect of Sox2 on transcription can also be seen when **SDm** is cotransfected with the Oct4 and Sox2 expression vectors into 293 cells (figure 5.1C). The more Sox2 vector is added to the cotransfection the higher the transcriptional activity becomes. Sox2 can bind to its binding site 34bp

upstream of the PORE, but not to the PORE. Thus, it helps activating transcription from its distal site, but does not compete for binding to the PORE, where the Oct4 dimer can display its potent transcriptional activity.

This interpretation is in agreement with all EMSAs shown in this chapter. The Sox2 can bind to the PORE, even without the distal Sox site, although it binds more efficiently when the distal PORE site is present (figure 5.2 cf. lanes 3 of DM and 5SDM). The higher stability of the Oct4/Sox2 heterodimer compared to the Oct4 homodimer on the PORE (figure 5.5) is also in agreement with the functional data: when Sox2 is expressed, transactivation of genes regulated by an SDM-type element, such as OPN, might have to be downregulated just in this manner – not ceasing expression altogether, but ensuring that it is reduced. Reduction of transcription is thus only achieved, if the less active heterodimer is more stable on the PORE than the more active homodimer. It is unclear at this stage, if another transcription factor or cofactor is required for further downregulation.

The question remains how Sox2 binds to the PORE. Fact is that the presence of the Sox site 34 bp upstream of the PORE somehow assists Sox2 recruitment to the PORE (figure 5.2cf. lanes 3 of DM and 5SDM; figures 5.5, 5.6 and 5.7). After additional analysis of the results I think looping or bending can be ruled out since, according to Irani *et al.* (1983), an integral number of helical turns of DNA has to be present between the two sites involved. This would be roughly the case for the Sox site and the PORE in the non-mutated SDM, which are 34bp apart. It could also be possible in the case of the 11SDM construct, in which the two sites are 11bp apart. In 5SDM though, there is only half a helical turn between the Sox site and the PORE. A number of EMSAs have been performed with this construct and the cotransfection experiments (figure 5.7) show that Sox2 has a negative effect on downstream gene

expression of 5SDM, as is the case for 11SDM and the non-mutated SDM. Thus, sliding or oozing appear more likely, as described and discussed in section 5.4.1 and 5.4.2 of the discussion in this chapter.

All results in this chapter were obtained by transfection and band shift assays. Certainly, there is no guarantee that Sox2 and Oct4 have the same effect on the endogenous *osteopontin* expression in cell lines and in the developing embryo. What points towards a net reduction of Sox2 on *osteopontin* expression, though, is the fact that the expression levels of Sox2 and *osteopontin* are inversely correlated. *Osteopontin* has a higher expression level in F9 than ES cells and Sox2 has a higher expression level in ES than F9 cells (figure 4.6 and Botquin *et al.*, 1998). Furthermore, the protein levels are also inversely correlated during the early phase of development. One feasible explanation for this inverse correlation is that Sox2 has a negative effect on *osteopontin* expression, thus the higher the Sox2 expression level the lower the *osteopontin* level would be, and vice versa. Net reduction, as discussed above, would be in agreement with this explanation, but would need to be further substantiated by alternative assays. One possible way would be to use a system, with which Sox2 could be inducibly expressed. This would allow to monitor the effect of Sox2 on the endogenous *osteopontin* gene and, even more interestingly, on stem cell fate.

Chapter 6

Conclusion

The work presented in this thesis supports earlier findings that transcription factors mainly act in combinatorial control with other transcription factors. This provides a means to integrate responses to a variety of signals using a limited number of transcription factors. More specifically my work focuses on how interaction surfaces on transcription factors increase the versatility of this family of proteins.

6.1 Transcription factors have multiple interaction surfaces

With the research I conducted and presented in this thesis I contributed to the concept of multiple interaction surfaces on transcription factors as an integrative approach to define the principles underlying differential complex formation on DNA (Tomilin *et al.*, 2000; Reményi *et al.*, 2001; Chapter 3 and Chapter 4 of this thesis). The structures of the Oct1 POU factor in complex with the MORE and PORE demonstrate how the same transcription factor can form unrelated arrangements using different, non-overlapping surface patches for domain-domain association (Reményi *et al.*, 2001). Each arrangement is induced by the specific positions and nature of the four protein segments binding to the respective DNA element. The resulting heterotetrameric arrangement of the POU dimers allows two alternative conformations of the POU molecules, one across the longitudinal center of each DNA response motif and one within each half-site. However, the crystal structures demonstrate that each complex reveals only one conformation, using two distinct interaction surfaces for the PORE and the MORE on both POU_S and POU_H. On the MORE the molecules bind across the center of the DNA element, whereas they each bind to one half-site on the PORE. In Chapter 3 I now propose that in the case of the POU1 dimer interacting with OBF1 on the PORE, the POU molecules are located

across the longitudinal center (similar to the MORE) but with the molecular interactions and subdomains being arranged as on the PORE.

Sox2 is another example of a transcription factor with more than one interaction surface (Chapter 4). Like for Oct1 different patches of the C-terminal part of the HMG domain interact with partner transcription factors. Oct1 as well as Sox2 depend on the specific DNA sequence to bring the interacting surface patch into position.

6.2 Interfaces interact with more than one type of transcription factor

These surface patches can interact with those of other members of their family as well as with surface patches of members of other transcription factor families. This results in a combinatorial variety that could enormously increase the complexity of the transcriptional network (Tomilin *et al.*, 2000; Reményi *et al.*, 2001; Chapter 3 and Chapter 4 of this thesis). For example, the POU_S interface, which can interact with the POU_H of the second molecule on the PORE can also interact with Sox2. Interestingly, when interacting with Sox2 the same interface of POU_S is in use on the *FGF4* element as on the *UTF1* promoter. The comparison of the *UTF1* sequence with that of the PORE and preliminary data suggest that Sox2/Oct4 heterodimerization on the PORE resembles that on *UTF1*. Furthermore, the interaction surface of Sox2 with Oct1 and Oct4 on *FGF4* seems to resemble that of Sox2 with Pax6 on the *DC5* element in the δ-crystallin promoter. Specifically, it subsumes that various combinations of transcription factors are possible, along with the potential of some of these proteins to interchange their quaternary DNA-mediated arrangements via one of multiple surfaces capable of protein-protein interactions (e.g. figure 4.7). Therefore, a

certain dimerization surface patch appears to be adept at mediating Velcro-like surface interactions with different interacting protein partners in a versatile fashion.

6.3 Surface patch interaction with cofactors

The interaction surface that is not being used in an Oct-dimer becomes exposed to the exterior of the protein (Reményi *et al.*, 2001). Thus the unused interface is accessible to other factors. As shown on the example of OBF1 binding to the Oct1 dimer on the PORE, cofactors can get recruited to these available surface patches (Chapter 3). On the PORE as well as on the octamer motif, OBF1 interacts with the same POU_S surface patch as POU_H does on the MORE. Since POU_H engrosses the POU_S surface patch, OBF1 cannot bind the MORE-type Oct1 dimer.

In addition to binding transcription factors, cofactors can also severely change the properties of that specific dimer. OBF1, for example, enhances Oct1 dimerization on PORE-like sequences. It loosens the sequence's stringency the heterotrimer can bind to, in terms of the sequence of the two Oct-binding sites and the spacing between them. Furthermore, OBF1 dramatically stabilizes the Oct1 dimer on the PORE and thus reduces its off-rate tremendously.

6.4 Different transcription factor combinations can bind to the same binding site

In addition to the fact that different transcription factors can bind to the same interaction surface, this may also happen on the same binding site. With the example of the PORE a binding element was revealed, on which Oct factors may homodimerize, Oct1 even with the cofactor OBF1, and also heterodimerize, as is the case for Oct4 and Sox2 (Botquin *et al.*, 1998; Chapter 5). Homodimers and

heterodimers compete for the same binding site. On the PORE, regulating the expression of *Osteopontin*, this results in activation of expression when the Oct4 homodimer is bound and repression of the Oct4-mediated activation when the Oct4/Sox2 heterodimer is recruited to the element. In this case it is specifically interesting to observe that an additional Sox binding site upstream of the PORE is required for Sox2 recruitment. The data presented in Chapter 5 are in agreement with the model that Sox2 is recruited to the PORE from the upstream Sox binding site by oozing, a mechanism that so far has only been postulated.

As such, the studies presented in this thesis provide insight into the adaptive mechanisms used by this set of transcription factors to assume a regulatory stronghold on various complex processes during mammalian development.

References

Ambrosetti, D.C., Basilico, C. and Dailey, L. (1997) Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol*, **17**, 6321-6329.

Ambrosetti, D.C., Schöler, H.R., Dailey, L. and Basilico, C. (2000) Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem*, **275**, 23387-23397.

Arulampalam, V., Eckhardt, L. and Pettersson, S. (1997) The enhancer shift: a model to explain the developmental control of IgH gene expression in B-lineage cells. *Immunol Today*, **18**, 549-554.

Assa-Munt, N., Mortishire-Smith, R.J., Aurora, R., Herr, W. and Wright, P.E. (1993) The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage lambda repressor DNA-binding domain. *Cell*, **73**, 193-205.

Aurora, R. and Herr, W. (1992) Segments of the POU domain influence one another's DNA-binding specificity. *Mol Cell Biol*, **12**, 455-467.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev*, **17**, 126-140.

Babb, R., Cleary, M.A. and Herr, W. (1997) OCA-B is a functional analog of VP16 but targets a separate surface of the Oct-1 POU domain. *Mol Cell Biol*, **17**, 7295-7305.

Basak, S. and Nagaraja, V. (1998) Transcriptional activator C protein-mediated unwinding of DNA as a possible mechanism for mom gene activation. *J Mol Biol*, **284**, 893-902.

Bergman, Y., Rice, D., Grosschedl, R. and Baltimore, D. (1984) Two regulatory elements for immunoglobulin kappa light chain gene expression. *Proc Natl Acad Sci USA*, **81**, 7041-7045.

Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C. and Kohwi-Shigematsu, T. (1992) Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science*, **255**, 195-197.

Bondurand, N., Pingault, V., Goerich, D.E., Lemort, N., Sock, E., Caignec, C.L., Wegner, M. and Goossens, M. (2000) Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. *Hum Mol Genet*, **9**, 1907-1917.

Botquin, V., Hess, H., Fuhrmann, G., Anastassiadis, C., Gross, M.K., Vriend, G., and Schöler, H.R. (1998) New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev*, **12**, 2073-2090.

Botquin, V., Lins, K., Reinbold, R., and Schöler, H.R. The Sox2-Oct4 heterodimer modulates PORE activity via a remote Sox2 regulatory element. *Submitted.*

Boulikas, T. (1993) Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human beta-globin gene complex. *J Cell Biochem*, **52**, 23-36.

Bowles, J., Schepers, G. and Koopman, P. (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev Biol*, **227**, 239-255.

Brehm, A., Ohbo, K. and Schöler, H. (1997) The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. *Mol Cell Biol*, **17**, 154-162.

Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*, **13**, 2465-2477.

Caelles, C., Hennemann, H. and Karin, M. (1995) M-phase-specific phosphorylation of the POU transcription factor GHF-1 by a cell cycle-regulated protein kinase inhibits DNA binding. *Mol Cell Biol*, **15**, 6694-6701.

Carey, M. (1998) The enhanceosome and transcriptional synergy. *Cell*, **92**, 5-8.

Casellas, R., Jankovic, M., Meyer, G., Gazumyan, A., Luo, Y., Roeder, R. and Nussenzweig, M. (2002) OcaB is required for normal transcription and V(D)J recombination of a subset of immunoglobulin kappa genes. *Cell*, **110**, 575-585.

Cepek, K.L., Chasman, D.I. and Sharp, P.A. (1996) Sequence-specific DNA binding of the B-cell-specific coactivator OCA-B. *Genes Dev*, **10**, 2079-2088.

Chasman, D., Cepek, K., Sharp, P.A. and Pabo, C.O. (1999) Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein-DNA interface. *Genes Dev*, **13**, 2650-2657.

Chi, N. and Epstein, J.A. (2002) Getting your Pax straight: Pax proteins in development and disease. *Trends Genet*, **18**, 41-47.

Cleary, M.A., Stern, S., Tanaka, M. and Herr, W. (1993) Differential positive control by Oct-1 and Oct-2: activation of a transcriptionally silent motif through Oct-1 and VP16 corecruitment. *Genes Dev*, **7**, 72-83.

Cleary, M.A. and Herr, W. (1995) Mechanisms for flexibility in DNA sequence recognition and VP16-induced complex formation by the Oct-1 POU domain. *Mol Cell Biol*, **15**, 2090-2100.

Clerc, R.G., Corcoran, L.M., LeBowitz, J.H., Baltimore, D. and Sharp, P.A. (1988) The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. *Genes Dev*, **2**, 1570-1581.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N. and Lovell-Badge, R. (1996) A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development*, **122**, 509-520.

Connor, F., Cary, P.D., Read, C.M., Preston, N.S., Driscoll, P.C., Denny, P., Crane-Robinson, C. and Ashworth, A. (1994) DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nucleic Acids Res*, **22**, 3339-3346.

Corcoran, L.M., Karvelas, M., Nossal, G.J., Ye, Z.S., Jacks, T. and Baltimore, D. (1993) Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. *Genes Dev*, **7**, 570-582.

Dailey, L., Yuan, H. and Basilico, C. (1994) Interaction between a novel F9-specific factor and octamer-binding proteins is required for cell-type-restricted activity of the fibroblast growth factor 4 enhancer. *Mol Cell Biol*, **14**, 7758-7769.

Dailey, L. and Basilico, C. (2001) Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. *J Cell Physiol*, **186**, 315-328.

Dawson, S.J., Yoon, S.O., Chikaraishi, D.M., Lillycrop, K.A. and Latchman, D.S. (1994) The Oct-2 transcription factor represses tyrosine hydroxylase expression via a heptamer TAATGARAT-like motif in the gene promoter. *Nucleic Acids Res*, **22**, 1023-1028.

De Santa Barbara, P., Bonneaud, N., Boizet, B., Desclozeaux, M., Moniot, B.,

Sudbeck, P., Scherer, G., Poulat, F. and Berta, P. (1998) Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Mullerian hormone gene. *Mol Cell Biol*, **18**, 6653-6665.

Delves, P.J. and Roitt, I.M. (2000a) The immune system. First of two parts. *N Engl J Med*, **343**, 37-49.

Delves, P.J. and Roitt, I.M. (2000b) The immune system. Second of two parts. *N Engl J Med*, **343**, 108-117.

Denhardt, D.T., Butler, W.T., Chambers, A.F. and Senger, D.R. (1995) Osteopontin: Role in Cell Signalling and Adhesion. *Ann N Y Acad Sci*, **760**.

Denny, P., Swift, S., Brand, N., Dabhade, N., Barton, P. and Ashworth, A. (1992) A conserved family of genes related to the testis determining gene, SRY. *Nucleic Acids Res*, **20**, 2887.

Falkner, F.G. and Zachau, H.G. (1984) Correct transcription of an immunoglobulin kappa gene requires an upstream fragment containing conserved sequence elements. *Nature*, **310**, 71-74.

Feldman, B., Poueymirou, W., Papaioannou, V.E., DeChiara, T.M. and Goldfarb, M. (1995) Requirement of FGF-4 for postimplantation mouse development. *Science*, **267**, 246-249.

Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R. and Bianchi, M.E. (1992) SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J*, **11**, 4497-4506.

Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. *Cell*, **55**, 757-769.

Frankel, A.D. and Kim, P.S. (1991) Modular structure of transcription factors:

implications for gene regulation. *Cell*, **65**, 717-719.

Frauwirth, K.A. and Thompson, C.B. (2002) Activation and inhibition of lymphocytes by costimulation. *J Clin Invest*, **109**, 295-299.

Fresno, M., Der Simonian, H., Nabel, G. and Cantor, H. (1982) Proteins synthesized by inducer T cells: evidence for a mitogenic peptide shared by inducer molecules that stimulate different cell types. *Cell*, **30**, 707-713.

Fukushima, A., Okuda, A., Nishimoto, M., Seki, N., Hori, T.A. and Muramatsu, M. (1998) Characterization of functional domains of an embryonic stem cell coactivator UTF1 which are conserved and essential for potentiation of ATF-2 activity. *J Biol Chem*, **273**, 25840-25849.

Giese, K., Cox, J. and Grosschedl, R. (1992) The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*, **69**, 185-195.

Gilbert, S.P. (2003) *Developmental Biology, 7th edition.* Sinauer Ass., Sunderland, MA

Goldfarb, M. (1996) Functions of fibroblast growth factors in vertebrate development. *Cytokine Growth Factor Rev*, **7**, 311-325.

Grindley, J.C., Hargett, L.K., Hill, R.E., Ross, A. and Hogan, B.L. (1997) Disruption of PAX6 function in mice homozygous for the Pax6Sey-1Neu mutation produces abnormalities in the early development and regionalization of the diencephalon. *Mech Dev*, **64**, 111-126.

Grosschedl, R., Giese, K. and Pagel, J. (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet*, **10**, 94-100.

Gstaiger, M., Knoepfel, L., Georgiev, O., Schaffner, W. and Hovens, C.M. (1995) A B-cell coactivator of octamer-binding transcription factors. *Nature*, **373**, 360-

362.

Gstaiger, M., Georgiev, O., van Leeuwen, H., van der Vliet, P. and Schaffner, W. (1996) The B cell coactivator Bob1 shows DNA sequence-dependent complex formation with Oct-1/Oct-2 factors, leading to differential promoter activation. *EMBO J*, **15**, 2781-2790.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, **346**, 245-250.

Harley, V.R., Jackson, D.I., Hextall, P.J., Hawkins, J.R., Berkovitz, G.D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P.N. (1992) DNA binding activity of recombinant SRY from normal males and XY females. *Science*, **255**, 453-456.

Hatzopoulos, A.K., Stoykova, A.S., Erselius, J.R., Goulding, M., Neuman, T. and Gruss, P. (1990) Structure and expression of the mouse Oct2a and Oct2b, two differentially spliced products of the same gene. *Development*, **109**, 349-362.

He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989) Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature*, **340**, 35-41.

Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G. and et al. (1988) The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. *Genes Dev*, **2**, 1513-1516.

Herr, W. and Cleary, M.A. (1995) The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev*, **9**, 1679-

1693.

Hinkley, C. and Perry, M. (1992) Histone H2B gene transcription during Xenopus early development requires functional cooperation between proteins bound to the CCAAT and octamer motifs. *Mol Cell Biol*, **12**, 4400-4411.

Hochschild, A. and Ptashne, M. (1986) Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell*, **44**, 681-687.

Ingraham, H.A., Chen, R.P., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell*, **55**, 519-529.

Ingraham, H.A., Flynn, S.E., Voss, J.W., Albert, V.R., Kapiloff, M.S., Wilson, L. and Rosenfeld, M.G. (1990) The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. *Cell*, **61**, 1021-1033.

Irani, M.H., Orosz, L. and Adhya, S. (1983) A control element within a structural gene: the gal operon of Escherichia coli. *Cell*, **32**, 783-788.

Jacobson, E.M., Li, P., Leon-del-Rio, A., Rosenfeld, M.G. and Aggarwal, A.K. (1997) Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev*, **11**, 198-212.

Jun, S. and Desplan, C. (1996) Cooperative interactions between paired domain and homeodomain. *Development*, **122**, 2639-2650.

Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R. and Kondoh, H. (1995) Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO J*, **14**, 3510-3519.

Kamachi, Y., Cheah, K.S. and Kondoh, H. (1999) Mechanism of regulatory target

selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. *Mol Cell Biol*, **19**, 107-120.

Kamachi, Y., Uchikawa, M. and Kondoh, H. (2000) Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet*, **16**, 182-187.

Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. and Kondoh, H. (2001) Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev*, **15**, 1272-1286.

Kapiloff, M.S., Farkash, Y., Wegner, M. and Rosenfeld, M.G. (1991) Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements. *Science*, **253**, 786-789.

Kelman, Z. and O'Donnell, M. (1995) Embryonic PCNA: a missing link? *Curr Biol*, **5**, 814.

Kim, U., Qin, X.F., Gong, S., Stevens, S., Luo, Y., Nussenzweig, M. and Roeder, R.G. (1996) The B-cell-specific transcription coactivator OCA-B/OBF-1/Bob-1 is essential for normal production of immunoglobulin isotypes. *Nature*, **383**, 542-547.

Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell*, **77**, 21-32.

Klemm, J.D. and Pabo, C.O. (1996) Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev*, **10**, 27-36.

Kong, X.P., Onrust, R., O'Donnell, M. and Kuriyan, J. (1992) Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, **69**, 425-437.

König, H., Pfisterer, P., Corcoran, L.M. and Wirth, T. (1995) Identification of CD36 as the first gene dependent on the B-cell differentiation factor Oct-2. *Genes Dev*, **9**, 1598-1607.

Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, **79**, 1233-1243.

Kuhlbrodt, K., Herbarth, B., Sock, E., Hermans-Borgmeyer, I. and Wegner, M. (1998a) Sox10, a novel transcriptional modulator in glial cells. *J Neurosci*, **18**, 237-250.

Kuhlbrodt, K., Herbarth, B., Sock, E., Enderich, J., Hermans-Borgmeyer, I. and Wegner, M. (1998b) Cooperative function of POU proteins and SOX proteins in glial cells. *J Biol Chem*, **273**, 16050-16057.

LaBella, F., Sive, H.L., Roeder, R.G. and Heintz, N. (1988) Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element. *Genes Dev*, **2**, 32-39.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Lang, D. and Epstein, J.A. (2003) Sox10 and Pax3 physically interact to mediate activation of a conserved c-RET enhancer. *Hum Mol Genet*, **12**, 937-945.

Laudet, V., Stehelin, D. and Clevers, H. (1993) Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res*, **21**, 2493-2501.

Lillycrop, K.A., Dent, C.L., Wheatley, S.C., Beech, M.N., Ninkina, N.N., Wood, J.N. and Latchman, D.S. (1991) The octamer-binding protein Oct-2 represses HSV immediate-early genes in cell lines derived from latently infectable sensory neurons. *Neuron*, **7**, 381-390.

Lin, Y.H., Huang, C.J., Chao, J.R., Chen, S.T., Lee, S.F., Yen, J.J. and Yang-Yen, H.F. (2000) Coupling of osteopontin and its cell surface receptor CD44 to the cell survival response elicited by interleukin-3 or granulocyte-macrophage colony-stimulating factor. [erratum appears in Mol Cell Biol 2001 Mar;21(6):2248.]. *Mol Cell Biol*, **20**, 2734-2742.

Lins, K., Reményi, A., Tomilin, A., Massa, S., Wilmanns, M., Matthias, P. and Schöler, H.R. (2003) OBF1 enhances transcriptional potential of Oct1. *EMBO J*, **22**, 2188-2198.

Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791-795.

Luo, Y., Fujii, H., Gerster, T. and Roeder, R.G. (1992) A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. *Cell*, **71**, 231-241.

Luo, Y. and Roeder, R.G. (1995) Cloning, functional characterization, and mechanism of action of the B- cell-specific transcriptional coactivator OCA-B. *Mol Cell Biol*, **15**, 4115-4124.

Maciejewski, P.M., Peterson, F.C., Anderson, P.J. and Brooks, C.L. (1995) Mutation of serine 90 to glutamic acid mimics phosphorylation of bovine prolactin. *J Biol Chem*, **270**, 27661-27665.

Maniatis, T., Falvo, J.V., Kim, T.H., Kim, T.K., Lin, C.H., Parekh, B.S. and Wathelet, M.G. (1998) Structure and function of the interferon-beta enhanceosome. *Cold Spring Harb Symp Quant Biol*, **63**, 609-620.

Mansouri, A., Hallonet, M. and Gruss, P. (1996) Pax genes and their roles in cell differentiation and development. *Curr Opin Cell Biol*, **8**, 851-857.

Marusic, A., Grcevic, D., Katavic, V., Kovacic, N., Lukic, I.K., Kalajzic, I. and Lorenzo, J.A. (2000) Role of B lymphocytes in new bone formation. *Lab Invest*, **80**, 1761-1774.

Matthias, P. (1998) Lymphoid-specific transcription mediated by the conserved octamer site: who is doing what? *Semin Immunol*, **10**, 155-163.

Melchers, F., Rolink, A., Grawunder, U., Winkler, T.H., Karasuyama, H., Ghia, P. and Andersson, J. (1995) Positive and negative selection events during B lymphopoiesis. *Curr Opin Immunol*, **7**, 214-227.

Mertin, S., McDowall, S.G. and Harley, V.R. (1999) The DNA-binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Res*, **27**, 1359-1364.

Miller, C.L., Feldhaus, A.L., Rooney, J.W., Rhodes, L.D., Sibley, C.H. and Singh, H. (1991) Regulation and a possible stage-specific function of Oct-2 during pre-B-cell differentiation. *Mol Cell Biol*, **11**, 4885-4894.

Minucci, S., Botquin, V., Yeom, Y.I., Dey, A., Sylvester, I., Zand, D.J., Ohbo, K., Ozato, K. and Schöler, H.R. (1996) Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy in vivo. *EMBO J*, **15**, 888-899.

Mittal, V., Cleary, M.A., Herr, W. and Hernandez, N. (1996) The Oct-1 POU-specific domain can stimulate small nuclear RNA gene transcription by stabilizing the basal transcription complex SNAPc. *Mol Cell Biol*, **16**, 1955-1965.

Murphy, F.V.t., Sweet, R.M. and Churchill, M.E. (1999) The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *EMBO J*, **18**, 6610-6618.

Nabel, G., Greenberger, J.S., Sakakeeny, M.A. and Cantor, H. (1981) Multiple

biologic activities of a cloned inducer T-cell population. *Proc Natl Acad Sci USA*, **78**, 1157-1161.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, **95**, 379-391.

Nielsen, P.J., Georgiev, O., Lorenz, B. and Schaffner, W. (1996) B lymphocytes are impaired in mice lacking the transcriptional co- activator Bob1/OCA-B/OBF1. *Eur J Immunol*, **26**, 3214-3218.

Nishimoto, M., Fukushima, A., Okuda, A. and Muramatsu, M. (1999) The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol*, **19**, 5453-5465.

Niswander, L. and Martin, G.R. (1992) Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development*, **114**, 755-768.

Niwa, H., Miyazaki, J. and Smith, A.G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet*, **24**, 372-376.

O'Regan, A. and Berman, J.S. (2000) Osteopontin: a key cytokine in cell-mediated and granulomatous inflammation. *Int J Exp Pathol*, **81**, 373-390.

Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Boon, K., Keaveney, M., Stunnenberg, H.G. and Muramatsu, M. (1998) UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *EMBO J*, **17**, 2019-2032.

Ovitt, C.E. and Schöler, H.R. (1998) The molecular biology of Oct-4 in the early mouse embryo. *Mol Hum Reprod*, **4**, 1021-1031.

Pabo, C.O. and Sauer, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem*, **61**, 1053-1095.

Palmieri, S.L., Peter, W., Hess, H. and Schöler, H.R. (1994) Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol*, **166**, 259-267.

Parslow, T.G., Blair, D.L., Murphy, W.J. and Granner, D.K. (1984) Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc Natl Acad Sci USA*, **81**, 2650-2654.

Pesce, M., Gross, M.K. and Schöler, H.R. (1998) In line with our ancestors: Oct-4 and the mammalian germ. *Bioessays*, **20**, 722-732.

Pesce, M., Anastassiadis, K. and Schöler, H.R. (1999) Oct-4: lessons of totipotency from embryonic stem cells. *Cells Tissues Organs*, **165**, 144-152.

Pesce, M. and Schöler, H.R. (2000) Oct-4: control of totipotency and germline determination. *Mol Reprod Dev*, **55**, 452-457.

Pevny, L.H. and Lovell-Badge, R. (1997) Sox genes find their feet. *Curr Opin Genet Dev*, **7**, 338-344.

Ptashne, M. (1986) Gene regulation by proteins acting nearby and at a distance. *Nature*, **322**, 697-701.

Qin, X.F., Reichlin, A., Luo, Y., Roeder, R.G. and Nussenzweig, M.C. (1998) OCA-B integrates B cell antigen receptor-, CD40L- and IL 4-mediated signals for the germinal center pathway of B cell development. *EMBO J*, **17**, 5066-5075.

Reim, G.M., T.; Stainier, D.Y., Kikuchi, Y. and Brand, M. (2004) The POU domain

protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. *Dev Cell*, **6**, 91-101.

Reményi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Schöler, H.R. and Wilmanns, M. (2001) Differential dimer activities of the transcription factor Oct-1 by DNA- induced interface swapping. *Mol Cell*, **8**, 569-580.

Reményi, A., Tomilin, A., Schöler, H.R. and Wilmanns, M. (2002) Differential activity by DNA-induced quarternary structures of POU transcription factors. *Biochem Pharmacol*, **64**, 979-984.

Reményi, A., Lins, K., Nissen, L.J., Reinbold, R., Schöler, H.R. and Wilmanns, M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev*, **17**, 2048-2059.

Rhee, J.M., Gruber, C.A., Brodie, T.B., Trieu, M. and Turner, E.E. (1998) Highly cooperative homodimerization is a conserved property of neural POU proteins. *J Biol Chem*, **273**, 34196-34205.

Roberts, S.B., Segil, N. and Heintz, N. (1991) Differential phosphorylation of the transcription factor Oct1 during the cell cycle. *Science*, **253**, 1022-1026.

Rolink, A. and Melchers, F. (1991) Molecular and cellular origins of B lymphocyte diversity. *Cell*, **66**, 1081-1094.

Ryan, A.K. and Rosenfeld, M.G. (1997) POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev*, **11**, 1207-1225.

Sadowski, I., Ma, J., Triezenberg, S. and Ptashne, M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, **335**, 563-564.

Sambrook, J., Fritsch E.F.and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.

Sauter, P. and Matthias, P. (1998) Coactivator OBF-1 makes selective contacts with both the POU-specific domain and the POU homeodomain and acts as a molecular clamp on DNA. *Mol Cell Biol*, **18**, 7397-7409.

Schöler, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N. and Gruss, P. (1989a) A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J*, **8**, 2543-2550.

Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N. and Gruss, P. (1989b) Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J*, **8**, 2551-2557.

Schöler, H.R., Dressler, G.R., Balling, R., Rohdewohld, H. and Gruss, P. (1990) Oct-4: a germline-specific transcription factor mapping to the mouse t- complex. *EMBO J*, **9**, 2185-2195.

Schoorlemmer, J. and Kruijer, W. (1991) Octamer-dependent regulation of the kFGF gene in embryonal carcinoma and embryonic stem cells. *Mech Dev*, **36**, 75-86.

Schubart, D.B., Rolink, A., Kosco-Vilbois, M.H., Botteri, F. and Matthias, P. (1996) B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. *Nature*, **383**, 538-542.

Schubart, D.B., Rolink, A., Schubart, K. and Matthias, P. (2000) Cutting edge: lack of peripheral B cells and severe agammaglobulinemia in mice simultaneously lacking Bruton's tyrosine kinase and the B cell- specific transcriptional coactivator OBF-1. *J Immunol*, **164**, 18-22.

Schubart, K., Massa, S., Schubart, D., Corcoran, L.M., Rolink, A.G. and Matthias, P. (2001) B cell development and immunoglobulin gene transcription in the absence of Oct-2 and OBF-1. *Nat Immunol*, **2**, 69-74.

Scully, K.M., Jacobson, E.M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D.W., Hooshmand, F., Aggarwal, A.K. and Rosenfeld, M.G. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science*, **290**, 1127-1131.

Segil, N., Roberts, S.B. and Heintz, N. (1991) Mitotic phosphorylation of the Oct-1 homeodomain and regulation of Oct- 1 DNA binding activity. *Science*, **254**, 1814-1816.

Soriano, N.S. and Russell, S. (1998) The Drosophila SOX-domain protein Dichaete is required for the development of the central nervous system midline. *Development*, **125**, 3989-3996.

Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A. and Baltimore, D. (1986) A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. *Nature*, **323**, 640-643.

Staudt, L.M., Clerc, R.G., Singh, H., LeBowitz, J.H., Sharp, P.A. and Baltimore, D. (1988) Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science*, **241**, 577-580.

Strubin, M., Newell, J.W. and Matthias, P. (1995) OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer- binding proteins. *Cell*, **80**, 497-506.

Struthers, R.S., Gaddy-Kurten, D. and Vale, W.W. (1992) Activin inhibits binding of transcription factor Pit-1 to the growth hormone promoter. *Proc Natl Acad Sci USA*, **89**, 11451-11455.

Stuart, E.T., Kioussi, C. and Gruss, P. (1994) Mammalian Pax genes. *Annu Rev Genet*, **28**, 219-236.

Sturm, R.A., Das, G. and Herr, W. (1988) The ubiquitous octamer-binding protein

Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev*, **2**, 1582-1599.

Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M. and Chambon, P. (1986) Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. *Nature*, **319**, 121-126.

Tinker-Kulberg, R.L., Fu, T.J., Geiduschek, E.P. and Kassavetis, G.A. (1996) A direct interaction between a DNA-tracking protein and a promoter recognition protein: implications for searching DNA sequence. *EMBO J*, **15**, 5032-5039.

Tokuzawa, Y., Kaiho, E., Maruyama, M., Takahashi, K., Mitsui, K., Maeda, M., Niwa, H. and Yamanaka, S. (2003) Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. *Mol Cell Biol*, **23**, 2699-2708.

Tomilin, A., Reményi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M. and Schöler, H.R. (2000) Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell*, **103**, 853-864.

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M. and Okuda, A. (2002) Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res*, **30**, 3202-3213.

Turque, N., Plaza, S., Radvanyi, F., Carriere, C. and Saule, S. (1994) Pax-QNR/Pax-6, a paired box- and homeobox-containing gene expressed in neurons, is also expressed in pancreatic endocrine cells. *Mol Endocrinol*, **8**, 929-938.

Underhill, D.A. and Gros, P. (1997) The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. *J Biol Chem*, **272**, 14175-14182.

Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., Van Leeuwen, H.C., Strating, M.J. and van der Vliet, P.C. (1992) The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J*, **11**, 4993-5003.

Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph*, **8**, 52-56, 29.

Walther, C. and Gruss, P. (1991) Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development*, **113**, 1435-1449.

Weber, G.F. and Cantor, H. (1996) The immunology of Eta-1/osteopontin. *Cytokine Growth Factor Rev*, **7**, 241-248.

Wegner, M. (1999) From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Res*, **27**, 1409-1420.

Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R., Laue, E.D. and Thomas, J.O. (1993) Structure of the HMG box motif in the B-domain of HMG1. *EMBO J*, **12**, 1311-1319.

Weiss, M.A. (2001) Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition. *Mol Endocrinol*, **15**, 353-362.

Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, **81**, 705-714.

Whitmarsh, A.J. and Davis, R.J. (2000) Regulation of transcription factor function by phosphorylation. *Cell Mol Life Sci*, **57**, 1172-1183.

Wu, T.J., Monokian, G., Mark, D.F. and Wobbe, C.R. (1994) Transcriptional activation by herpes simplex virus type 1 VP16 in vitro and its inhibition by oligopeptides. *Mol Cell Biol*, **14**, 3484-3493.

Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999)

Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, **13**, 1263-1275.

Yang, J., Muller-Immergluck, M.M., Seipel, K., Janson, L., Westin, G., Schaffner, W. and Pettersson, U. (1991) Both Oct-1 and Oct-2A contain domains which can activate the ubiquitously expressed U2 snRNA genes. *EMBO J*, **10**, 2291-2296.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Schöler, H.R. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development*, **122**, 881-894.

Yuan, H., Corbi, N., Basilico, C. and Dailey, L. (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev*, **9**, 2635-2645.